

UNIVERZITA KARLOVA V PRAZE

FILOZOFICKÁ FAKULTA

Katedra psychologie



Diplomová práce

**KONEKCIONISTICKÉ MODELOVÁNÍ VÝVOJE NĚKTERÝCH
KOGNITIVNÍCH PROCESŮ SPOJENÝCH S OSVOJOVÁNÍM
JAZYKOVÉ KOMPETENCE V ČESKÉM JAZYCE**

**CONNECTIONIST MODELLING OF DEVELOPMENT OF SOME COGNITIVE
PROCESSES ASSOCIATED WITH ACQUISITION OF LANGUAGE SKILLS IN
CZECH LANGUAGE**

Vypracoval: Luděk Stehlík

Vedoucí diplomové práce: PhDr. Pavel Uhlář

Praha

2007

Prohlašuji, že jsem diplomovou práci vypracoval samostatně a že jsem uvedl všechny použité prameny a literaturu.

V Praze, dne 28. června 2007

.....

Chtěl bych poděkovat vedoucímu diplomové práce PhDr. Pavlovi Uhlářovi za poskytnuté rady a připomínky a především pak za jeho dlouhodobou podporu, kterou mi poskytoval v mém teoreticko-výzkumném počínání v průběhu celého mého studia.

Obsah

Summary	viii
Resumé	x
Úvod	xii
OBEČNÁ ČÁST	1
1 Kognitivistické paradigma	1
1.1 Mentalistický diskurz	1
1.2 Mysl jako výpočet	13
1.2.1 Formální systémy	13
1.2.1.1 Interpretované formální systémy	14
1.2.1.2 Historický a kognitivně antropologický kontext formálních systémů ..	19
1.2.1.3 Formální logika	42
1.2.2 Automatické formální systémy	48
1.2.2.1 Turingův stroj	53
1.2.2.2 Počítače, von Neumannovská architektura a virtuální stroje.....	60
1.2.3 Výpočetní teorie mysli	68
1.2.3.1 Sémanticky transparentní systémy	69
1.2.3.2 Symbolové systémy	71
1.2.3.3 Konekcionismus	84
1.2.3.4 Výpočetní architektura mysli/mozku	96
1.2.3.5 Distribuované kognitivní systémy	107
2 Konekcionistické modelování kognitivních procesů	124
2.1 Principy zpracovávání informací v lidském mozku	124
2.2 Konekcionistické modely	133
2.2.1 Matematické modelování dynamických systémů	134
2.2.1.1 Teorie dynamických systémů	135
2.2.1.2 Základní nástroje vizualizace a analýzy chování komplexních dynamických systémů	153
2.2.1.3 Některé další nástroje zkoumání komplexních dynamických systémů	158
2.2.1.3.1 Příčinné smyčkové diagramy	159

2.2.1.3.2 Kognitivní mapy	161
2.2.1.3.3 Dynamické modelování	165
2.2.1.3.4 Fuzzy kognitivní mapy	167
2.2.1.4 Dynamická kognitivní věda	187
2.2.2 Matematické modelování dynamiky šíření a zpracovávání informací v lidském mozku	213
2.2.2.1 Základní pojmy	214
2.2.2.2 Automatické učící algoritmy	220
2.2.2.2.1 Hebbovo pravidlo	221
2.2.2.2.2 Delta pravidlo	225
2.2.2.2.3 Perceptronové konvergenční pravidlo	226
2.2.2.2.4 Pravidlo nejmenších čtverců	226
2.2.2.2.5 Učící algoritmus zpětného šíření	231
2.2.2.2.6 Genetické algoritmy	234
2.2.2.3 Vliv architektury neuronové sítě na způsob zpracovávání informací... ..	240
2.2.2.3.1 Vícevrstevné neuronové sítě	240
2.2.2.3.2 Jednoduché rekurentní sítě	246
2.2.2.3.3 Atraktorové sítě	248
2.2.2.3.4 Kompetitivní sítě	254
2.2.2.3.5 Autoasociátory	257
2.2.2.3.6 Bias node	260
2.3 Metodologie konekcionistických simulací kognitivních procesů.....	261
2.3.1 Tréninková fáze	265
2.3.2 Hodnotící fáze	268
2.3.3 Analytická fáze	271
3 Vývojová psycholingvistika	281
3.1 Základní charakteristiky jazyka	282
3.1.1 Hláskový systém jazyka	284
3.1.2 Morfologický systém jazyka	286
3.1.3 Lexikální systém jazyka	286
3.1.4 Syntaktický systém jazyka	287
3.1.5 Diskurz	290

3.1.6 Prozodie	291
3.2 Osvojování jazyka	291
3.2.1 Protosociální chování dítěte, intuitivní rodičovství a „maminkovština“	292
3.2.2 Fáze osvojování jazyka	294
3.2.3 Vývoj některých vybraných kognitivních procesů spojených s osvojováním jazyka	297
3.2.3.1 Senzitivita k prozodické struktuře jazyka	297
3.2.3.2 Senzitivita k hláskové struktuře jazyka	298
3.2.3.3 Segmentace spojitého toku řeči	299
3.2.3.4 Učení se významům slov	302
3.2.3.5 Osvojování gramatických pravidel	305
3.2.3.5.1 Sémantický bootstrapping	307
3.2.3.5.2 Syntaktický bootstrapping	308
3.2.3.5.3 Distribuční analýza	309
3.2.3.5.4 Konekcionistický model distribuční analýzy	310
3.2.3.5.5 Empirický výzkum osvojování gramatiky	321
3.2.4 Přínos konekcionistického modelování pro výzkum (nejen) jazykového vývoje	322
3.2.4.1 Nativismus vs. empirismus	323
3.2.4.2 Kontinuální vs. stupňovitý vývoj	333
3.2.4.3 Způsob reprezentace poznatků	340
3.2.4.4 Kritická období	341
3.2.4.5 Operacionalizace vývojově-psychologických pojmů	341
SPECIÁLNÍ ČÁST	343
4 PDP model využití fonotaktických a prozodických vodítek k segmentaci spojitého toku řeči na jednotlivá slova v ČJ	343
4.1 Výzkumný problém	343
4.2 Definování úlohy pro neuronovou síť	345
4.3 Funkční architektura neuronové sítě	348
4.4 Tréninkové parametry neuronové sítě (experimentální design)	349
4.5 Některé předběžné výsledky	352

4.5.1 Nová definice hranice slov	356
4.5.2 Formulace hypotéz	359
4.6 Výsledky a diskuse	364
4.7 Závěr	376
Seznam literatury	379
Internetové zdroje	392
Seznam doprovodného materiálu na přiloženém CD	393
Příloha	394

Summary

The thesis discusses questions about the nature of the human mind, the process of language acquisition and the cognitive development in general and presents results of connectionist modelling of one of the seminal developmentally-psycholinguistic phenomena – segmentation of continuous speech flow on the basis of phonotactic and prosodic cues. In the general part of the thesis there is at first briefly introduced the philosophical topic of mind-body problem and then there is described intellectual, historical and cognitively-anthropological background of the computational-representational paradigm which is in the contemporary cognitive science the most prevailing view of the human mind treating the mental life as a kind of computation in the sense of manipulation and transformation of syntactic structures. There are in detail introduced two competing interpretations of this statement in the form of physical symbol system hypothesis and connectionism which differ primarily about the opinion on how far from the level of ordinary thought and concepts we should look for the formal rules specifying mental computation. There are also mentioned two alternative theories in the form of dynamic (or embodied) cognitive science and theory of distributed cognition according to which the nature of intelligence doesn't lie (only) in the manipulation of complex inner representations of the outer world but mainly in dynamical interaction between larger number of various factors occurring inside and outside the organism. What follows is a short introduction to the theory of dynamical systems with explanation of concepts like phase space, attractor, bifurcation and the range of other concepts and tools which are used for description of dynamical systems' behavior, including the behavior of the range of various types of artificial neural networks. In much more detail there are discussed biological and technical principles behind the working of connectionist models and the methodology of connectionist modeling of cognitive processes. Subsequently there are introduced some basic characteristics of language and the main developmental tasks faced by child acquiring its mother language like speech segmentation, acquiring a meaning of words and grammar. In the context of the nature vs. nurture debate there is described the connectionist approach to some of these topics which offers utterly new perspective not only on the nature of innate predispositions which enable relatively effortless and rapid language acquisition but also on the nature of the very cognitive development and the knowledge (mental representations) acquired by a child. Special part of the thesis refers to research project within which was with the help of *Tlearn* simulation software created connectionist model of segmentation of

continuous speech flow using the statistical information about the distribution of sounds in the selected sample of Czech language and intonation cues contained in the speech in the form of regular stress on the first syllable of a word. Elman's Simple Recurrent Network (SRN) was used and it was trained on uninterrupted succession of sounds forming several randomly ordered words in diverse grammatical forms. The task of the net was to anticipate its next input in two different experimental conditions – with and without the information about the location of stress on the first syllable of certain word. The results of the first simulation experiment confirmed that SRN is able to discover statistical regularities in cooccurrence of sounds in language input which is demonstrated in differential erroneous performance of SRN in the task of prediction of various sounds in dependence on their location in a word. Thanks to this effect word boundaries are in continuous speech flow clearly marked as locations where the conditional probability of the next sound, given the preceding sounds, is very low. The usage of inflecting language demonstrated that similar principle can also be used to locate the boundary between the root of a word and its grammatical ending. By reason of possibility to distinguish a root of a word apart from a grammatical ending a new definition of a word boundary was proposed as a location of considerable contrast between conditional probability of first sound and conditional probability of one or more immediately following sounds. The results of the first simulation experiment demonstrated that addition of prosodic cue into training environment of UNS significantly ($p < 0,0001$) sharpens the contrast defining the word boundary in the continuous speech flow and by that it facilitates its identification. Further analysis revealed that this effect is caused partly by worse performance of SRN in the task of anticipation of the first sound of a word and partly by better performance of SRN in the task of anticipation of one or more sounds which follow immediately. The second simulation experiment revealed that the second mechanism could be only a consequence of unrepresentative content of the training environment and that if we used more representative sample of Czech language in which in the position tightly behind the first sound with the stress would occur as many different sounds as in the position tightly behind the same sound but without stress the contribution of prosodic cue for speech segmentation would be mediated mainly (or exclusively) by more erroneous performance of UNS in the task of anticipation of the first sound of a word. The results of realized simulations indicate that Elman's SRN in association with newly defined word boundary is plausible model of word segmentation which is able to implement within one single conceptual framework the usage of phonotactic and prosodic cues which are assumed to be exploited in some way by small children in the process of speech segmentation.

Resumé

Diplomová práce diskutuje otázky spojené s povahou lidské mysli, procesu osvojování jazyka a kognitivního vývoje a prezentuje výsledky konekcionistického modelování jednoho z klíčových vývojově psycholingvistických jevů – segmentace spojitého toku řeči na jednotlivá slova za využití fonotaktických a prozodických vodítek. V obecné části práce je nejdříve uvedeno filozofické téma psychofyzického problému a následně je popsáno myšlenkové, historické a kognitivně-antropologické pozadí v současné kognitivní vědě zcela převládajícího komputačně-reprezentačního paradigmatu, v rámci kterého je lidská mysl pojímána jako druh výpočtu ve smyslu manipulace a transformace syntaktických struktur. Podrobně jsou představeny dvě soupeřící interpretace tohoto tvrzení v podobě hypotézy fyzikálního symbolového systému a konekcionismu, které se liší zejména v názoru na to, jak blízko (nebo daleko) běžným myšlenkám a pojmům by se měla hledat formální pravidla specifikující mentální výpočet. Zmíněny jsou rovněž dvě alternativní teorie v podobě dynamické (nebo také tělesné) kognitivní vědy a teorie distribuované kognice, podle kterých podstata inteligence nespočívá (pouze) v manipulaci komplexními vnitřními reprezentacemi okolního světa, ale především v dynamické interakci mezi větším počtem různých faktorů, které se nacházejí jak uvnitř tak i vně organismu. Následuje stručný úvod do problematiky (matematického) modelování dynamických systémů, kde jsou vysvětleny pojmy jako fázový prostor, atraktory, bifurkace a řada dalších pojmů a nástrojů, které se používají k popisu chování dynamických systémů, včetně chování řady různých druhů umělých neuronových sítí. Podrobněji jsou v práci pojednány biologické a technické principy, na nichž je založena činnost konekcionistických modelů, a také metodologie spojená s konekcionistickým modelováním kognitivních procesů. Následně jsou v práci stručně představeny některé základní charakteristiky jazyka a z nich vyplývající hlavní vývojové úkoly, které stojí před dítětem osvojujícím si svůj mateřský jazyk, jako je segmentace řeči, osvojování významu slov a gramatiky jazyka. V kontextu debaty nativismus vs. empirismus je pak popsán přístup k některým z těchto témat z perspektivy konekcionismu a jeho specifického pojetí lidské mysli jako statistického (a paralelně-distribuovaného) procesoru, které nabízí zcela nový pohled nejen na povahu vrozených předpokladů, které dítěti umožňují se snadno a rychle naučit svému mateřskému jazyku, ale také na povahu samotného kognitivního vývoje a dítětem osvojovaných poznatků (mentálních reprezentací). Speciální část práce pak referuje o výzkumném projektu, v jehož rámci byl s pomocí simulačního softwaru *Tlearn* vytvořen

konekcionistický model segmentace spojitého toku řeči za využití statistických informací o distribuci jednotlivých hlásek ve vybraném vzorku českého jazyka a intonačních nápořád obsažených v řeči v podobě stálého přízvuku na první slabice slova. Použita byla Elmanova jednoduchá rekurentní síť (SRN), která byla trénována na nepřerušované posloupnosti hlásek tvořících několik náhodně seřazených slov v různých gramatických tvarech. Úloha sítě spočívala v anticipaci svého příštího vstupu ve dvou různých experimentálních podmínkách - s a bez informace o umístění stálého přízvuku na první slabice určitého slova. Výsledky prvního simulačního experimentu potvrdily, že SRN dokáže v jazykovém vstupu objevit statistické zákonitosti v souvškytu jednotlivých hlásek, které se projevují v její rozdílné chybovosti při predikci jednotlivých hlásek podle toho, na jaké pozici ve slově se tyto hlásky nacházejí, díky čemuž jsou ve spojitém toku řeči jasné vyznačeny hranice jednotlivých slov jako místa nízké podmíněné pravděpodobnosti výskytu další hlásky za předpokladu výskytu všech předcházejících hlásek. Použití flektivního jazyka prokázalo, že obdobný princip je možné využít rovněž při lokalizaci hranic mezi kořenem slova a jeho gramatickou koncovkou. Z důvodu možnosti odlišení začátku slova od gramatické koncovky byla navržena nová definice hranice slova jako místa výrazného kontrastu mezi podmíněnou pravděpodobností výskytu první hlásky a jedné nebo více bezprostředně následujících hlásek. Výsledky prvního simulačního experimentu prokázaly, že přidání prozodického vodítka do tréninkového prostředí UNS statisticky signifikantně ($p < 0,0001$) vyostřuje kontrast, který definuje hranici slova ve spojitém toku řeči, a tak i usnadňuje její identifikaci. Další analýza prokázala, že tento efekt je způsoben jednak zvýšením chybovosti UNS při predikci první hlásky a jednak snížením chybovosti UNS při anticipaci jedné nebo více hlásek, které po první hlásce bezprostředně následují. Druhý simulační experiment demonstroval, že v případě druhého mechanismu by se mohlo jednat pouze o artefakt nereprezentativního obsahu tréninkového prostředí UNS a že při použití reprezentativnějšího vzorku českého jazyka, ve kterém by se na pozici těsně za první hláskou s přízvukem prostřídalo stejné množství různých hlásek jako za stejnou hláskou bez přízvuku, by byl přínos prozodického vodítka pro segmentaci spojitého toku řeči zprostředkován především (nebo výhradně) nárůstem chybovosti UNS při anticipaci první hlásky nového slova. Celkově výsledky provedených simulací naznačují, že Elmanova jednoduchá rekurentní síť ve spojení s nově definovanou hranicí slov představuje plausibilní model segmentace řeči, který v rámci jednoho konceptuálního rámce dokáže implementovat proces exploatace fonotaktického a prozodického vodítka, o kterých se předpokládá, že je malé děti v nějaké podobě využívají při segmentaci spojitého toku řeči.

Úvod

„Skutečná záhada: „Já“ hledí dovnitř a nenachází nic než neurony.“

PAUL BROKS, *Do země ticha. Lidský mozek – tajemný svět „uvnitř hlavy“*

„Sì, abbiamo un' anima. Ma è fatta di tanti piccoli robot.“¹

GIULIO GIORELLI

„Dvě věci naplňují mysl obdivem a úctou tím většími, čím častěji a vytrvaleji o nich přemýšlíme;“² lidská mysl ve mně a jazyk, kterým mohu nejrůznější její obsahy přenášet do mysli druhých lidí. Tato volná parafráze slavného Kantova výroku výstižně vyjadřuje nejen hlavní předmět této diplomové práce, který spadá do oblasti vývojové a kognitivní psychologie a psycholingvistiky, ale také autorovu vlastní motivaci k jejímu sepsání. Lidská mysl a jazyk byly odedávna považovány za dva výsostné znaky člověka, které ho měly odlišovat od zbytku živočišné říše, a jako takové vždy fascinovaly a současně provokovaly zvědavost a představivost různých myslitelů, kteří se snažili přijít na kloub tomu, co stojí v základech těchto dvou specificky lidských duševních mohutností. Museli se při tom potýkat s absencí vhodné sady dostatečně jemných a zároveň dostatečně přesných pojmových nástrojů, s jejichž pomocí by bylo možné se zmocnit podstaty jinak velice těžko uchopitelných duševních jevů s jejich nehmotnou, neviditelnou a vysoce abstraktní povahou (jakkoli závislou na lidského mozku jakožto biologickém substrátu lidského myšlení). Pojmové nástroje, kterými tito myslitelé disponovali, byly příliš hrubé na to, aby se jim podařilo lidskou mysl a s ní spojené duševní schopnosti bezezbytku spoutat zákony tohoto světa a vysvětlit, jak by je bylo možné fyzicky realizovat.

Tato situace se změnila až někdy okolo poloviny 20. století, kdy se díky pokrokům v teorii formálních systémů objevila „nová úroveň analýzy, která je nezávislá na fyzice, ale přesto je ve své podstatě mechanistická... věda o struktuře a funkci oddělené od materiální substance“³ (Pylyshyn, 1986, s. 68). Tato nová úroveň analýzy umožnila vznik kognitivní vědy - nové vědy o kognitivních procesech lidské mysli, která začala být považována za formu výpočtu ve smyslu manipulace a transformace syntaktických struktur. Právě této základní myšlenky kognitivní vědy se detailně věnuje první kapitola diplomové práce, jejíž obsah má do značné

¹ „Ano, máme duši. Ta se ale skládá z velkého množství drobných robotů.“ Citováno podle Dennett, 2004a, s. 1

² Citováno podle Rádl, 1999, s. 301

³ „...a new level of analysis, independent of physics yet mechanistic in spirit... a science of structure and function divorced from material substance.“

míry charakter jakési kognitivní archeologie či kognitivní antropologie, která odkrývá myšlenkové, ale i sociální a historické pozadí počítačně-reprezentačního paradigmatu, které má v současné kognitivní vědě naprosto dominantní postavení, jakkoli se v posledních deseti dvaceti letech objevují určité formy „hereze“ v podobě dynamické (nebo také tělesné) kognitivní vědy či teorie distribuované kognice, které jsou zde rovněž stručně představeny.

Kromě prezentace (a určité kritiky) některých základních principů kognitivistického paradigmatu jsou v rámci první kapitoly uvedeny a na několika málo konkrétních příkladech také demonstrovány dvě hlavní interpretace tohoto specifického pojetí lidské mysli a z nich vyplývající možné přístupy k modelování kognitivních procesů prostřednictvím symbolových a konekcionistických výpočetních modelů. Největší pozornost je přitom věnována modelům založeným na paralelně distribuovaném zpracovávání informací – tzv. PDP (parallel distributed processing) či konekcionistickým modelům, známým též pod názvem umělé neuronové sítě. Těžiště tohoto tématu se nachází ve druhé kapitole. Vzhledem k vysoce dynamické povaze chování řady různých druhů PDP systémů je toto téma uvedeno stručným úvodem do problematiky (matematického) modelování dynamických systémů a s tím spjaté terminologie, na kterou je v průběhu textu práce často odkazováno. Samostatně jsou pak pojednány technické a biologické principy, na nichž je založena činnost PDP systémů, které matematicky modelují dynamiku šíření a zpracovávání informací v lidském mozku. Poměrně detailně je popsána rovněž metodologie konekcionistického modelování kognitivních procesů.

Díky svému vysoce specifickému způsobu zpracovávání informací a schopnosti se učit na základě svých vlastních „zkušeností“ PDP modely zásadním způsobem ovlivnily probíhající debatu o povaze mentálních reprezentací, kognitivního (a speciálně jazykového) vývoje a za ním se skrývajících vrozených predispozic umožňujících dítěti například se snadno a rychle naučit svému mateřskému jazyku. Právě tato problematika je předmětem třetí kapitoly. Ta se nejdříve věnuje několika základním charakteristikám jazyka, od kterých potom plynule přechází k některým hlavním tématům vývojové psycholingvistiky jako jsou fáze osvojování jazyka, řeč zaměřená na dítě (tzv. *baby-talk* nebo také *child-directed speech*), segmentace plynulého proudu řeči na jednotlivá slova, senzitivita dítěte k prozodické a hláskové struktuře jazyka, osvojování významu slov a gramatických pravidel atp. V závěru této kapitoly jsou nastíněny způsoby přístupu k některým z těchto vývojově psycholingvistickým témat z perspektivy konekcionismu a je zde také poukázáno na možný přínos konekcionistického modelování pro výzkum jazykového a obecně kognitivního vývoje.

Ten je pak demonstrován ve čtvrté a poslední kapitole, která referuje o výzkumné části diplomové práce, v jejímž rámci byl s pomocí simulačního softwaru *Tlearn* vytvořen konekcionistický model jednoho z klíčových vývojově psycholingvistických jevů, který se objevuje v průběhu osvojování jazykové kompetence – model segmentace spojitého toku řeči na základní jednotky (slova) za využití statistických informací o distribuci jednotlivých hlásek ve vybraném vzorku českého jazyka a intonačních nápodůb obsažených v řeči v podobě stálého přízvuku na první slabice slova. Výstupem této simulace je - kromě ilustrace konkrétní aplikace konekcionistické metodologie, která není v české odborné psychologické obci příliš často využívána – hlubší vhled do (možné) povahy mentálních procesů a reprezentací souvisejících s osvojováním jazykové kompetence v českém jazyce.

Vzhledem k vysoce abstraktní povaze řady myšlenek a pojmů uváděných v této práci jsem se při jejich popisu a vysvětlování vždy snažil o co možná největší názornost, a to prostřednictvím různých metafor či přirovnání a četných obrazových ilustrací (což bohužel také přispělo k výraznému nárůstu počtu stránek). Práce je proto místy koncipována téměř jako didaktická příručka – čtenář v ní nalezne četné odkazy na webové stránky a na doprovodný materiál umístěný na přiloženém CD, díky čemuž má čtenář jedinečnou možnost si řadu zde uváděných myšlenek a pojmů „osahat“ a vyzkoušet tak říkajíc na vlastní kůži. Navzdory značnému rozsahu práce (a jejímu mnohoslibnému názvu) není tato práce ničím jiným než nepatrným škrábnutím po povrchu ohromného množství poznatků nashromážděných za bezmála 50 let existence kognitivně psychologického a psycholingvistického výzkumu, a jako taková tak v žádném případě nemůže být považována za nějaký vyčerpávající přehled všech teorií, hypotéz či výzkumů, které kdy byly vysloveny nebo provedeny v rámci studia osvojování jazyka. Její přínos – vedle nepatrného rozšíření poznatků o možnostech jednoho konkrétního konekcionistického modelu jednoho konkrétního psycholingvistického jevu – spočívá především v inspiraci pro další výzkum vedený v intencích konekcionistického pohledu na podstatu lidské mysli. Na závěr tohoto úvodu mohu už jen vyslovit přání, aby čtenář při čtení této diplomové práce zažil stejné intelektuální dobrodružství, jaké zažíval její autor při její realizaci.

OBECNÁ ČÁST

1 Kognitivistické paradigma

„In our daily lives we all predict and explain other people's behavior from what we think they know and what we think they want. Beliefs and desires are the explanatory tools of our own intuitive psychology, and intuitive psychology is still the most useful and complete science of behavior there is. To predict the vast majority of human acts – going to refrigerator, getting on the bus, reaching into one's wallet – you don't need to crank through a mathematical model, run a computer simulation of a neural network, or hire a professional psychologist; you can just ask your grandmother.“

STEVEN PINKER, *How The Mind Works*

„How can intelligence emerge from nonintelligence? ... you can build a mind from many little parts, each mindless by itself. I call „Society of Mind“ this scheme in which each mind is made of many smaller processes. These we'll call agents. Each mental agent by itself can only do some simple thing that needs no mind or thought at all. Yet when we join these agents in societies – in certain very special ways – this leads to true intelligence .“

MARVIN MINSKY, *The Society of Mind*

„Sedl si na pohovku, která se zdála býti skutečnou, nebo pokud skutečná nebyla, tak ho alespoň udržela, a vzhledem k tomu, že právě to se od pohovek očekává, tak to podle všech relevantních hledisek skutečná pohovka byla.“

DOUGLAS ADAMS, *Život, vesmír a vůbec*

1.1 Mentalistický diskurz

Behaviorismus, jak známo, ve své klasické podobě zcela ignoroval vnitřní obsahy a procesy lidského myšlení a činnost psychologie omezoval pouze na studium objektivně pozorovatelných vztahů mezi souborem podnětů a reakcí (Hoskovec, Nakonečný, Sedláková, 2002). Období nadvlády tohoto přístupu v psychologii je často líčeno jako „doba temna“, která byla zažehnána „světlem rozumu“ s příchodem kognitivní revoluce na přelomu 50. a 60. let 20. století. Jakkoli se tento pohled na věc (při srovnání explanační a predikční úspěšnosti behavioristických a kognitivistických teorií) přímo nabízí, je dobré si uvědomit, že všichni lidé jsou v určitém slova smyslu behavioristy. Z druhého člověka je nám vždy dostupné pouze jeho námi smyslově zaznamenané chování, a na vnitřní duševní stavy a procesy tak můžeme pouze nepřímo usuzovat. Tuto naši situaci dobře vystihuje „stará čínská vyprávěnka o dvou mudrcích, kteří stáli na mostě přes bystřinu a sledovali vyskakující ryby. Jeden z nich prohlásil, že ryby se ze studené a proudící vody radují. Druhý na to reagoval argumentem, jak

že to může vědět, není-li sám rybou. První mu opáčil, co může vědět o jeho vědomostech a vnitřním životě, není-li jím samým.“ (Komárek, 2000, s. 28) Přes všechna tato naše omezení se však dokážeme ve svých každodenních životech celkem úspěšně „nabourávat“ do černé skříňky skrytých psychických stavů a procesů, které jsou podkladem vnějších behaviorálních projevů druhých lidí. Dovedeme totiž bez nějakého většího úsilí a vědomé námahy nahlížet sebe sama a jiné lidské bytosti (a nejen je) jako zvláštní druh entit, které se pohybují samy od sebe, vnímají své okolí, sledují nějaké cíle, mají nějaké touhy, záměry, přání a přesvědčení. To nám umožňuje „číst“ duševní stavy lidí okolo nás, jakoby nahlížet, co se jim odehrává v jejich myslích, a tak i předpovídat jejich budoucí chování.

Naše uvažování o chování druhých lidí takto vždy probíhá v rámci tzv. **mentalistického diskurzu** (také *folková* či *laická psychologie*, *mentalizace*, *teorie mysli* nebo *propozičně-postojová psychologie*), který se při vysvětlování a predikci lidského chování odvolává na různé mentální entity jako jsou přesvědčení, přání, potřeby, obavy apod. (Sedláková, 2004). Typickou podobu těchto vysvětlení zachycuje termín *propozičně-postojová psychologie* (*propositional attitude psychology*) - takové vysvětlení propojuje určitý druh mentálního postoje (jako je „věřit“, „domnívat se“, „doufat“, „obávat se“ apod.) s určitým druhem tvrzení („že venku prší“, „že v ledničce je vychlazené pivo“ apod.) (Clark, 2001). Tímto způsobem lze například velice úspěšně vysvětlit to, že někdo směřuje k ledničce, tím, že daná osoba se chce napít a že se tato osoba také domnívá, že v ledničce je vychlazené pivo, kterým by mohla svou žízeň uhasit. U zdravých dětí¹ se tento specifický způsob uvažování o chování druhých lidí objevuje kolem čtvrtého roku věku², ale dospělému člověku již „naskakuje“ jaksí sám od sebe a automaticky a v běžných situacích je pro člověka de facto „neviditelný“ a nikomu nepřijde na mysl, že by bylo možné uvažovat i jiným způsobem. Kdybychom chtěli parafrázovat *Ludwiga Wittgensteina*, mohli bychom říci, že se jedná o jistý druh iluze, která

¹ V důsledku abnormálního vývoje mozku nebo jeho poškození při úrazu nebo cévní mozkové příhodě může dojít k poškození schopnosti mentalizace, která je podle dosavadních funkčně-zobrazovacích studií vázána – vedle dalších mozkových oblastí - především na pravostrannou mediální prefrontální kůru. Podle některých výzkumů lze vysvětlit typické příznaky dětského autismu (výrazné deficity v komunikaci, v sociální interakci, ve schopnosti vytvářet citové a sociální vztahy a ve schopnosti hrát různé druhy her) jako vývojovou poruchu funkčního modulu mentalizace neboli teorie mysli (Koukolík, 2002, 2006).

² Při výzkumu vývoje teorie mysli u malých dětí se často používá nějaká forma loutkového testu jako je například *test Sally-Anne*: Sally a Anne jsou loutky; dítě sleduje Sally, která je v místnosti, jak dává do košíčku kostku. Potom Sally z místnosti odchází a na scénu přichází Anne, která vezme kostku z košíčku a přemístí ji a schová do krabičky. Na otázku, kde bude Sally hledat kostku, odpovídají děti mladší čtyř let, že ji bude hledat v krabičce. Taková odpověď naznačuje, že takto staré děti si ještě plně neuvědomují, že každý člověk se nachází v určitém specifickém vnitřním duševním stavu, který je odrazem jeho interakcí s okolním světem. Důsledkem je to, že ve výše popsané experimentální situaci nejsou schopny jasně rozlišit mezi tím, co vědí ony samy a co ví Sally, a uvědomit si tak tedy, že Sally reprezentuje svůj svět mylně. Děti starší čtyř let to již obvykle dokáží.

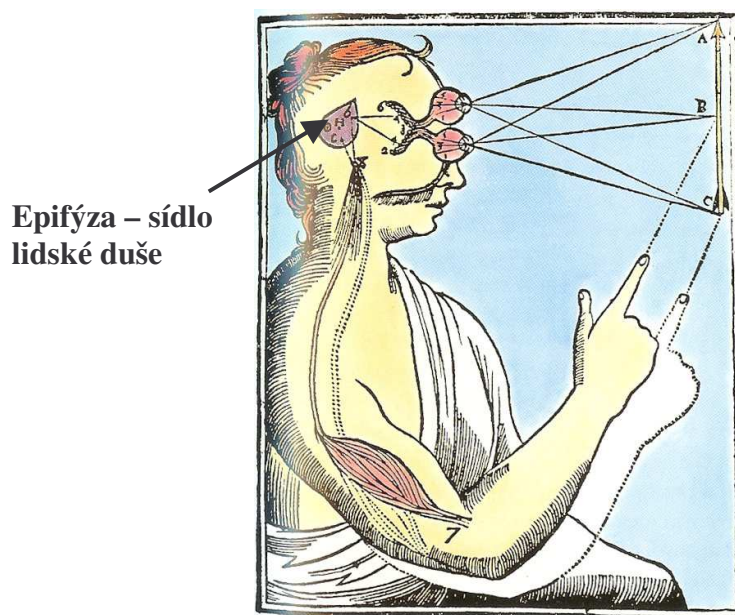
člověku sedí jako brýle na nose, takže cokoli pak vidí, vidí skrze ni, a tak ho tato iluze drží v zajetí (Wittgenstein, 1993, s. 62, 65).

Explicace chování lidí v rámci mentalistického diskurzu umožňuje člověku generovat relativně přesná očekávání a předpovědi ohledně budoucího chování druhých lidí. Jestliže budu předpokládat, že někdo nechce zmoknout, a dále, že je přesvědčen, že účinným způsobem, jak v dešti nezmoknout, je vzít si deštník, potom na základě těchto předpokladů mohu předvídat, že jestliže bude ráno pršet, vezme si tento člověk s sebou deštník. Tento druh predikce hraje klíčovou roli de facto ve všech našich složitějších životních projektech. Každou chvíli například plánujeme svou činnost na základě něčího tvrzení, že se objeví v určitý den, v určitý čas a na určitém místě; takové tvrzení odráží určitou intenci či záměr (že se objeví v určitý den, v určitý čas a na určitém místě), který kauzálně ovlivňuje chování člověka, který toto tvrzení pronesl - takto můžeme naplánovat setkání s člověkem i několik měsíců dopředu s přesností na centimetry a na minuty.

Je otázka, jak je možné, že nám pojmání lidí jako mentálních agentů disponujících nejrozličnějšími mentálními stavy umožňuje tak úspěšně vysvětlovat a predikovat jejich chování. Koneckonců nikdo nikdy žádný mentální stav na vlastní oči neviděl. Jediné, k čemu máme přístup, jsou vnější projevy těchto mentálních stavů. Folková psychologie, jak se zdá, na základě viditelných vnějších projevů postuluje existenci jakýchsi „neviditelných“ mentálních entit, které spolu vstupují do řady vzájemných vztahů, které mají povahu zákonů typu „jestliže někdo chce X a je přitom přesvědčen, že k jejímu získání musí udělat Y, potom udělá Y“. Folkovou psychologii lze tedy považovat za jistý druh teorie o skrytých a neviditelných, přitom ale kauzálně působících příčinách inteligentního chování, která celkem úspěšně vysvětluje zákonité řetězení různých mentálních stavů a behaviorálních projevů (Clark, 2001). Co jsou ale ony mentální stavy? Na tuto otázku byly v průběhu dějin psychologického myšlení dávány v zásadě dva základní typy odpovědí.

Jedna odpověď spočívala v postulování existence jakési nemateriální substance, která byla sídlem všech duševních stavů a dějů a která přitom zároveň dokázala kauzálně působit na hmotné tělo, tak aby se toto chovalo v souladu s intencemi duševní substance. Nejznámějším příkladem tohoto přístupu je *Descartova* dualistická koncepce lidské mysli jako na těle v zásadě nezávislé entity, *res cogitans* (věc poznávající), která přesto dokáže kauzálně ovlivňovat chování lidského těla, které je jakožto *res extensa* (věc rozprostraněná) součástí hmotného světa, který se řídí svými vlastními pravidly a zákony, které jsou odlišné od těch, kterými se řídí duševní svět lidské subjektivity (Störig, 1993). Hlavní námitka vůči dualistickým koncepcím vyplývá z jejich základního předpokladu, kterým je postulování

existence dvou zcela rozdílných substancí, které přitom ale spolu musí nějakým způsobem interagovat (tzv. **mind-body problem** nebo také **psychofyzický problém**). Na jedné straně musí tělesné smyslové orgány doručit lidské mysli smyslové počítky a vjemy a na straně druhé musí lidská mysl, poté, co všechno náležitě promyslí, prostřednictvím těla vyprodukovat reakci, která by byla adekvátní tomu, co se právě dozvěděla prostřednictvím svých smyslů - tzv. **karteziánský interakcionismus** či **interakcionistický dualismus** (Dennett, 1993). Budeme-li ale chtít udržet základní předpoklad dualismu, totiž že lidská mysl a tělo představují dvě zcela zásadně odlišné substance, je téměř nemožné mezi nimi dosáhnout jakékoli smysluplné interakce. V rámci karteziánského interakcionismu se tento problém nakonec podařilo „vyřešit“ pouze za cenu přijetí velice krkolomného předpokladu v podobě **predeterminovaného** (psychofyzického) **paralelismu**, kdy je činnost lidské mysli a těla koordinována průběžnými božími zásahy; v této souvislosti se také hovoří o tzv. *okkasionalismu*, od latinského *occasio*, zde ve smyslu *příležitosti* volního aktu, kdy bůh zajišťuje jeho projev ve fyzické činnosti lidského těla (Störig, 1993).



Obrázek 1: Ilustrace Descartovy teorie mysli. Obraz šípů je přes oči přenesen do šišinky (epifýzy), sídla lidské duše, která přenesený obraz zpracuje a na základě svých vlastních úmyslů pak prostřednictvím „duchů“ (spiritus) vyšle povel svalům pravé ruky, tak aby tato ukázala požadovaným směrem.

Z důvodu této vnitřní rozpornosti dualistických koncepcí se filozofové a vědci při vysvětlování povahy mentálních jevů uchylují spíše k **monistickým koncepcím**, které postulují existenci pouze jednoho druhu substance, která má většinou materiální povahu. Před

materialisty přitom stojí dosti obtížný úkol, jak vysvětlit vysoce specifické vlastnosti mentálních jevů v pojmech, které by byly plně kompatibilní s explanacemi chování ne-mentálního světa hmoty. Mentální svět se vyznačuje především dvěma charakteristikami: 1) Jednotlivé mentální stavy (myšlenky, pocity či přání) dávají vyvstat dalším mentálním stavům a behaviorálním projevům ve specifickém kauzálním řetězci, který respektuje to, co bychom mohli nazvat „rozumové důvody“, „rozum“ či „racionalita“; například myšlenka, že dnes bude pršet, způsobí, že si vezmu deštník, obléknu si nepromokavou bundu a pomyslím si „...zase dneska bude ošklivě a zase si nezhraju tenis“. 2) Druhým rysem mentálního života je jistý kvalitativní pocit (tzv. *qualia*), který je spojen s mentálními stavy a procesy; příkladem by mohl být pocit červenosti červené barvy na přebalu knihy nebo pocit chtivosti spojený s touhou po něčem apod. Materialista musí všechny tyto vlastnosti mentálního života uvést do souvislosti se strukturou, organizací a kauzálním působením materiálního světa. Podle mnohých představitelů kognitivní vědy již disponujeme teorií, která to dokáže – alespoň co se týče prvního rysu mentálního života. Tato teorie, tzv. **výpočetní teorie mysli** (*computational theory of mind*), která bude podrobněji popsána v následujícím oddíle, je schopna vysvětlit racionální řetězení mentálních stavů a vzorců chování čistě na základě strukturálních vlastností systému, ve kterém se myšlení odehrává. Oblíbeným objektem, který se obvykle používá jako ilustrace základní myšlenky výpočetní teorie mysli, je jednoduchá kapesní elektronická kalkulačka, která nám dá na výstupu součet dvou čísel poté, co na klávesnici vyťukáme posloupnost nějakých tlačítek, která běžně interpretujeme například jako „5“ „+“ „3“ a „=“. Aby kalkulačka něco takového dokázala, nepotřebuje nic vědět o číslech ani o jednotlivých aritmetických operacích. Kalkulačka je jednoduše zkonstruována tak (tj. má takovou fyzikální a kauzální strukturu), aby stlačení dané posloupnosti tlačítek vždy vedlo k výstupu „8“ stejně jistě, jako kámen hozený do vzduchu padá po co nejkratší možné dráze směrem k zemi. Jediný rozdíl mezi kalkulačkou a kamenem je v tom, že kalkulačka byla navržena tak, aby celkové chování jejích jednotlivých součástí odráželo aritmetická pravidla a omezení, kterými se řídí přechody mezi čísly poté, co jsou na ně aplikovány aritmetické operace. Podobně je pak struktura našeho mozku uspořádána (evolucí či učením) tak, aby jeho fyzikální fungování odráželo rozumová pravidla a omezení, kterými se řídí posloupnosti přechodů mezi jednotlivými mentálními stavy a vzorci chování. Předpověď deštivého počasí pak může vést k vyhledání deštníku ve skříni nebo k nostalgickému vzpomínání na léto stejně snadno jako vyťukání příslušných tlačítek na kalkulačce vede k výslednému součtu čísel. Obě je záležitost pouze struktury fyzikálního systému, jehož chování určitým způsobem odráží

pravidla a zákonitosti dané domény, kterou může být aritmetika, šachy nebo zdravý selský rozum. Stručně řečeno, „*mysl je pouze šikovně uspořádaná hmota*“¹ (Clark, 2001, s. 5).

Tímto se dostáváme k důležité otázce, kterou je vztah mezi teorií folkové psychologie, která je tak úspěšná při explanaci a predikci lidského chování, a vědeckou (výpočetní) teorií lidské mysli, jejíž obrysy již bylo možné zahlédnout ze stručného expozé v předchozím odstavci. Podle Clarka (2001) existují v zásadě tři možnosti, jak se k této otázce postavit. Prominentním představitelem prvního přístupu je kognitivní vědec *Jerry Fodor*, podle kterého je teorie folkové psychologie při explanaci a predikci chování lidí tak úspěšná proto, že je zkrátka pravdivá, tzn. že jí postulované mentální entity mají svůj reálný protějšek v kognitivním systému, konkrétně v jeho vnitřních fyzikálních stavech, které mohou kauzálně působit. Jinými slovy, podle Fodora se obsahy a struktura mentalistického diskurzu a vnitřní fyzikální stavy kognitivního systému dokonale překrývají (Fodor, 1987). Takže jestliže budu predikovat, že si nějaký člověk vezme deštník, protože jsem se nějakým způsobem dozvěděl, že je tento člověk přesvědčen o tom, že dnes bude pršet, potom podle Fodora úspěch mé predikce závisí na tom, že mozek tohoto člověka obsahuje nějaký vnitřní fyzikální stav, který lze interpretovat jako „dnes bude pršet“, a dále že tento vnitřní stav způsobí vznik dalších mentálních stavů (například „bylo by dobré si vzít deštník“ nebo „...kam mi děti zase zašantročily můj deštník!?“) a určitou behaviorální reakci (půjde a vezme si deštník ze skříně). Teoretickou implementací tohoto přístupu je **reprezentační teorie mysli**, resp. **hypotéza fyzikálního symbolového systému**: propoziční postoje nejsou ničím jiným než vnitřními reprezentacemi a ty jsou zase vnitřními stavy nějakého fyzikálního kognitivního systému; mentální procesy jsou pak ve své fyzikální podstatě kauzálními procesy, které způsobují („logice“ duševního života člověka odpovídající) přechody mezi jednotlivými vnitřními reprezentacemi (podrobněji viz následující oddíl).

Zcela opačný přístup než Fodor zastává neurofilozof *Paul Churchland* (1981), podle kterého mentální entity postulované folkovou psychologií nemají ve vnitřních stavech lidského kognitivního systému (tj. mozku) žádný jasně identifikovatelný korelát. Churchland, na rozdíl od Fodora, považuje teorii folkové psychologie pouze za povrchní, zkreslený a nepřesný popis toho, co se děje v lidském kognitivním systému, což je podle Churchlanda také důvod toho, proč je mentalistický diskurz úspěšně aplikován pouze v několika málo oblastech a ve spoustě dalších (jako je např. kreativita, paměť, spánek, duševní nemoci, senzomotorická koordinace v reálném čase atp.) k našemu porozumění duševním jevům nijak

¹ „*mindfulness is just a matter, nicely orchestrated.*“

zvláště nepřispívá. Teoretickým ztělesněním tohoto přístupu je **konekcionismus**, v jehož výpočetních modelech inspirovaných kognitivní architekturou lidského mozku s distribuovanými reprezentacemi je jen velmi obtížné, ba skoro nemožné najít jasné koreláty obsahů a struktury mentalistického diskurzu (více viz oddíl „1.2.3.3 Konekcionismus“).

Třetí přístup k otázce vztahu folkové psychologie a vědecké teorie mysli reprezentuje kognitivní vědec a filozof *Daniel Dennett*, podle kterého sice pravděpodobně neexistuje těsná korespondence mezi obsahy a strukturou mentalistického diskurzu a vnitřními stavy fyzikálního kognitivního systému, to ale podle něj nikterak neubírá folkovou psychologii na její explanační adekvátnosti. Je to obdoba situace, kdy na atomární úrovni nedokážeme najít koreláty tekutosti nebo mokrosti vody, přesto je na určité úrovni popisu velice užitečné a praktické mluvit o vodě jako o tekuté a mokré substanci. Podle Dennetta (1987) jsou mentální entity, se kterými operuje folková psychologie, abstrakta, která jsou vyvozena ze vzorců inteligentního chování určitých druhů objektů, a to nezávisle na konkrétní povaze příčin, které toto chování generují. V této souvislosti Dennett hovoří o zaujímání tzv. **intencionálního postoje**.

Podle Dennetta (2004b) existují tři základní druhy postojů ke světu, které představují tři základní strategie předpovídání. **Fyzikální postoj** využívá ke generování předpovědí o chování objektů to, co víme o fyzikálních zákonech a fyzikálních vlastnostech předmětů, o jejichž budoucí chování nám jde. Fyzikální postoj používáme například tehdy, když očekáváme, že kámen puštěný z ruky dopadne na zem. Abychom dokázali předpovědět budoucí chování kamene, to jediné, co musíme udělat, je připsat mu hmotnost, resp. váhu a svou předpověď založit na znalosti gravitačního zákona. Vedle kamenů a dalších podobných objektů existuje skupina vyprojektovaných objektů, v jejichž případě lidé většinou používají jiný způsob předpovídání – ten vychází z **projektového postoje**, který je (ve vztahu k fyzikálnímu postoji) určitou zkratkou: Takto například od někoho dostaneme nový model digitálního budíku, se kterým vůbec neumíme zacházet a nevíme, co od něj máme očekávat. Po zevrubném prozkoumání jeho tlačítek a displeje však brzy zjistíme, že *když* určitá konkrétní tlačítka určitým konkrétním způsobem zmáčkne a na displeji se nám ukáží určité konkrétní číselné údaje, *potom* za několik hodin budík vydá zvuk dostatečně silný na to, aby nás vzbudil a umožnil nám přijít včas do práce, do školy, na schůzku apod. Když se snažíme porozumět budíku, a predikovat tak jeho chování, nepotřebujeme složitě a namáhavě odvozovat jednotlivé fyzikální zákony umožňující tento typ pravidelného chování. Nemusíme budík rozmontovat, změřit a zvážit jeho jednotlivé části, určit elektrické napětí v jeho

obvodech. Stačí nám předpokládat, že budík je určitým způsobem vyprojektován – právě jako budík – a že se podle tohoto projektu bude i chovat.

Vedle kamenů, digitálních budíků a dalších podobných objektů existuje jistá speciální třída objektů, vůči kterým obvykle zaujímáme **intencionální postoj**, který spočívá v nahlížení daného objektu jako **racionálního konatele** či **agenta**, který volbu svého chování podřizuje úvahám o svých **přesvědčeních** a **přáních**. Tento postoj můžeme (trochu antropomorficky) aplikovat rovněž na náš digitální budík – takový budík se pak pro nás stane jakýmsi osobním časoměřičem, který bezchybně a bez odmlouvání plní naše instrukce: Když mu nařídíme, aby nás vzbudil, tak, že mu dáme na srozuměnou přesný čas, kdy nás má vzbudit, můžeme se spolehnout na jeho schopnost zjistit, kdy tento námi stanovený čas nastal, a splnit to, co od nás dostal přikázáno (tedy probudit nás). Skrze brýle intencionálního postoje se nám pak bude zdát, že takový budík nabude přesvědčení, že právě nyní nastal onen námi stanovený čas, kdy nás má vzbudit, a že díky našim instrukcím bude mít budík motivaci podle těchto instrukcí jednat. K tomu, abychom porozuměli, proč budík dělá to, co dělá, tento antropomorfismus nutný není, ale v případě artefaktů složitějších než je budík je zaujmutí intencionálního postoje užitečnější a někdy i nezbytné. Příkladem může být počítačový šachový program. Ať už jsou fyzikální nebo projektové charakteristiky počítače, na němž šachový program běží, jakékoli, lze jeho chování velice úspěšně interpretovat zaujmutím intencionálního postoje: Takový počítač si představujeme jako racionálního agenta či konatele, který **chce** vyhrát a který **zná** pravidla šachu a aktuální rozestavení figurek na šachovnici. Z intencionálního postoje je interpretace chování počítače mnohem snazší než z postoje fyzikálního či projektového. Jak uvádí Peregrin (2003), by bylo samozřejmě teoreticky možné popsat činnost šachového programu jako posloupnost změn aktivace milionů prvků čipu počítače, ke kterým dochází, když tento program běží v počítači, ale v takovém popisu by se nikdo nedokázal (v rozumném čase) zorientovat, a jako takový by tedy byl naprosto k ničemu. Šachovému programu či počítači, na němž šachový program běží, lze skutečně porozumět jedině tak, že budeme jeho chování interpretovat v termínech strategií, které používá a v jejichž důsledku určitým způsobem hraje. V kterémkoli okamžiku šachové hry se můžeme podívat na rozestavení figurek na šachovnici a udělat si seznam všech přípustných tahů, které může počítač (v souladu s pravidly šachu) udělat, když je na tahu. Při předpovídání dalšího tahu počítače nám postačí brát v úvahu pouze přípustné tahy, protože můžeme předpokládat, že počítač, resp. jeho šachový program je racionálním agentem, jehož **cílem** (přáním) je šachovou partii vyhrát a který **ví**, že chce-li vyhrát, musí dělat pouze a jedině přípustné tahy, kterých se tedy, jakožto racionální konatel, bude držet. Naši předpověď jeho chování můžeme

ještě zpřesnit, jestliže všechny tyto přípustné tahy ohodnotíme a seřadíme z hlediska jejich možného příspěvku k budoucí výhře (či prohře) od těch nejlepších k těm nejhorším. Na základě tohoto žebříčku pak můžeme předpovědět, že počítač provede ten tah, který je pro něj nejvýhodnější, neboť už je v přirozené povaze všech racionálních agentů, že vyhledávají své vlastní dobro a vyhýbají se zlu. Podstatou intencionálního postoje je tak tedy to, že abychom mohli předpovídat a explikovat chování nějakého konatele, musíme přijmout předpoklad, že tento konatel není hloupým konatelem, který by mohl udělat jakoukoli hloupost, ale konatelem chytrým, který bude jednat chytře z hlediska své omezené perspektivy, kterou popisujeme tak, že tomuto konateli připisujeme různá specifická přesvědčení a přání, která jsou založena na jeho vnímání situace a na jeho cílech a potřebách. Takový konatel (*intencionální systém*) je tedy informačně (reprezentačně) modulovanou a cíle vyhledávající entitou, entitou, která racionálně sleduje své cíle prostřednictvím toho, že jedná způsoby, které jí určuje její vnímání okolností. A právě tyto vlastnosti (jakkoli „hardwarově“ realizované) jsou tím, co umožňuje intencionální postoj využívat k předpovídání a vysvětlování chování intencionálních systémů (Dennett, 2004b).

Dennetovi se často vytýká, že klade rozhodnutí o připsání mentálních stavů a procesů nějaké entitě čistě do rukou pozorovatele. Dennett (1987) však zdůrazňuje, že toto připisování se děje na základě objektivně existujících charakteristik chování, které jsou na pozorovateli zcela nezávislé. Tyto charakteristiky jsou však v chování rozpoznatelné pouze z perspektivy intencionálního postoje, který určité (objektivně existující) charakteristiky chování zvýrazní a jiné zase upozadí, a to způsobem, jehož konečným výsledkem je obraz racionálního agenta, který se chová na základě svých přání a přesvědčení. Při zaujetí ne-mentalistického postoje člověk tyto zákonitosti a vzorce v chování nezachytí. Clark (2001) to přirovnává k situaci, kdy máme dvě různé číselné řady (například „2, 4, 6, 8, 10“ a „50, 52, 54, 56, 58“), které sdílejí jednu společnou, objektivně existující zákonitost (+2), kterou však můžeme odhalit pouze z perspektivy „aritmického postoje“. Intencionální postoj je podobnou perspektivou, aplikovanou ovšem na chování lidí, resp. na chování všech entit, které mají ten správný „design“, aby mohli být nazírány jako intencionální systémy.

Jakkoli by se mohlo zdát, že Dennettův přístup je jakýmsi kompromisem mezi Fodorovým a Churchlandovým přístupem (neboť Dennett na jedné straně uznává, stejně jako Fodor, explanační adekvátnost folkové psychologie a na straně druhé spolu s Churchlandem popírá existenci nějakého vnitřního neurálního kódu, který by těsně korespondoval s obsahy a strukturou mentalistického diskurzu), ve skutečnosti je Dennettova argumentace k argumentacím Fodora a Churchlanda mimoběžná. Zatímco Fodor a Churchland sdílejí

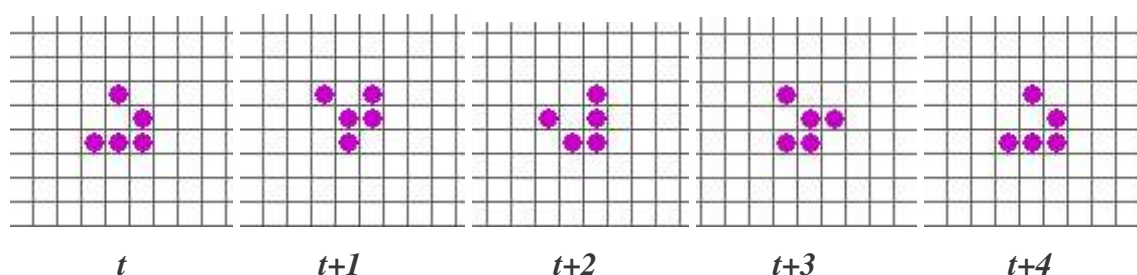
společný předpoklad, že teorie folkové psychologie vyžaduje ospravedlnění v podobě prokázané existence těsných neurálních korelátů mentálních stavů a procesů, podle Dennetta žádný takový korelát není k prokázání užitečnosti mentalistického diskurzu zapotřebí. Dennett (1998) tento svůj přístup ilustruje na příkladě jednoduchého *celulárního automatu* (dále CA) - diskrétního dynamického systému, jehož stav se mění v diskrétních časových krocích. Svět CA je většinou tvořen pravidelnou strukturou buněk v N -rozměrném prostoru; nejčastěji $N = 2$; potom hovoříme o tzv. 2D CA, jehož struktura má podobu pravidelné čtvercové mřížky. Každá buňka v mřížce přitom může nabývat K různých stavů; většinou se jedná pouze o dva stavy: zapnuto/vypnuto, „živá“ buňka/„mrtvá“ buňka, 1/0 apod. Hodnota stavu buňky v každém následujícím časovém kroku se počítá podle tzv. lokální přechodové funkce, která počítá nový stav buňky na základě jejího aktuálního stavu v kombinaci s aktuálním stavem sousedních buněk. Výpočet nových hodnot stavů všech buněk přitom probíhá paralelně. (Mařík, Štěpánková, Lažanský a kol., 2001) Dennett pro svou ilustraci použil „Hru života“ (*Game of Life*) od matematika *Johna Conwaye*.¹ Jedná se o známý a oblíbený 2D CA, v němž se aktivace šíří mřížkou podle tří následujících pravidel, která se aplikují na (maximálně) osm sousedních buněk každé buňky:

- Pravidlo 1: Jestliže právě dvě sousední buňky jsou aktivní (1), stav buňky, ať už je jakýkoli (1/0), se nemění.
- Pravidlo 2: Jestliže právě tři sousední buňky jsou aktivní (1), buňka se stává aktivní (1), ať už je její aktuální stav jakýkoli (1/0).
- Pravidlo 3: V jakémkoli jiném případě se buňka stává neaktivní (0), ať už je její aktuální stav jakýkoli (1/0).

Navzdory jednoduchým pravidlům, kterými se řídí celý „vesmír“ Conwayova CA, poté, co je do mřížky zasetá určitá počáteční aktivita (v podobě aktivace několika málo buněk), můžeme se stát svědky vzniku (emergence) velmi zajímavých vzorců chování. Jednotlivé buňky se shlukují do nejrůznějších tvarů a obrazců; tyto obrazce mohou zanikat, neměně trvat, cyklicky se opakovat nebo se posunovat v 2D prostoru. Co je však podle Dennetta nejdůležitější, tyto vynořující se vzorce chování můžeme popisovat pomocí slovníku, který operuje na vyšší úrovni abstrakce než je slovník „fyzikálních“ pravidel Conwayova „vesmíru“. Takto můžeme například mluvit o „kluzáku“ (*glider*), jestliže se na určitém místě v mřížce CA objeví vzorec

¹ Ten, kdo by chtěl blíže prozkoumat vlastnosti 2D univerza Conwayova CA, může tak učinit s pomocí dvou počítačových programů (*Winlife* a *Life32*), které jsou obsaženy na přiloženém CD v adresáři *2D_CA*.

aktivace, který bude vypadat jako ten na prvním políčku obrázku 2; aplikací výše uvedených pravidel se tento vzorec aktivace v průběhu čtyř časových kroků vrátí přes řadu čtyř různých mezilehlých obrazců do své výchozí podoby a přitom se zároveň v 2D prostoru trochu posune směrem šikmo doprava; a v tomto svém putování bude „kluzák“ pokračovat do té doby, dokud neukončíme simulaci, nebo dokud se mu do cesty nepostaví některá z dalších aktivních buněk ve mřížce CA.

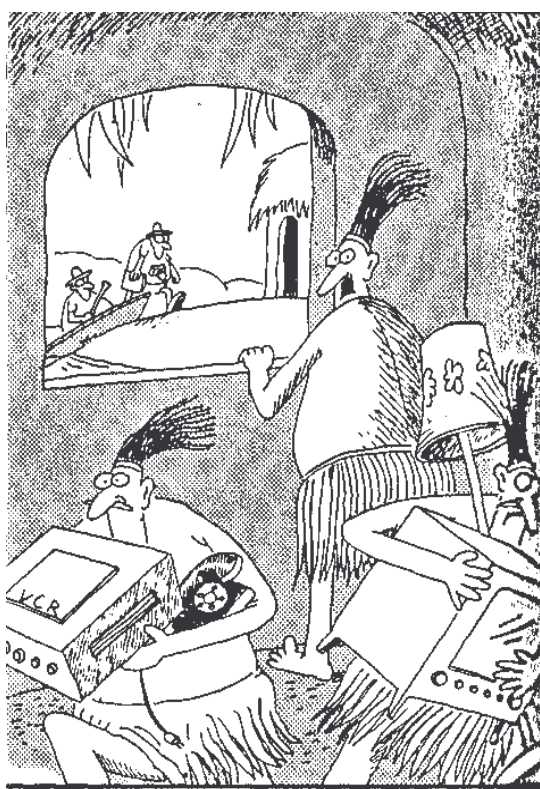


Obrázek 2: Posloupnost změn aktivačního vzorce v mřížce 2D CA, kterou lze na abstraktnější úrovni popisu charakterizovat jako „kluzák“, který si v průběhu času zachovává svou „identitu“ a přitom se dokáže pohybovat 2D prostorem Conwayova vesmíru. V terminologii teorie dynamických systémů se jedná o druh limitního cyklu, kdy se určitá podmnožina buněk v mřížce CA po stejném počtu kroků vrací do stejného stavu. (vytvořeno s pomocí softwaru Winlife)

Podle Dennetta diskurz odkazující k entitám typu „kluzák“ (nebo „blikač“, „had“, „bochník“, „ropucha“, „jedlík“, „maják“ a mnoho dalších, na které lze v Conwayově CA běžně narazit) identifikuje v Conwayově vesmíru objektivně existující vzorce a zákonitosti, které vnějšímu pozorovateli umožňují vytvářet různé užitečné generalizace a předpovědi ohledně chování tohoto 2D světa. Přestože tedy v posledku stojí v pozadí všech těchto aktivačních vzorců „fyzikální“ zákony Conwayova CA, bylo by podle Dennetta velkou chybou ignorovat explanačně velice užitečnou úroveň popisu, která odkazuje ke „kluzákům“ nebo „blikačům“.

To, co platí pro svět Conwayova CA, platí rovněž pro lidské chování. Kognitivní vědec Zenon Pylyshyn (1986) jako ilustraci užitečnosti té úrovně popisu, která pracuje s vysokoúrovňovými mentalistickými koncepty, uvádí příběh člověka, který se nejdříve stane svědkem dopravní nehody, poté běží k telefonní budce a nakonec vyťuká číslo 911 (číslo na pohotovost v USA). Podle Pylyshyna se samozřejmě můžeme pokusit vysvětlit chování svědka dopravní nehody prostřednictvím nějaké fyzikální teorie, která pracuje s neurony, neurotransmitery, ionty apod., avšak takové vysvětlení nám nijak nepomůže v porozumění rozumovým důvodům, které stojí v pozadí chování takového člověka. Čistě fyzikální teorie

by nám například nevnesla ani trochu světla do situace, kdy by telefon přestal fungovat, nebo kdy by se nehoda nestala v USA, ale ve Velké Británii. Čistě fyzikální popis jednotlivých vzorců chování by se totiž významně lišil podle toho, zda by náš svědek dopravní nehody musel vyťukat číslo 999 (číslo na pohotovost ve Velké Británii), nebo číslo 911 (číslo na pohotovost v USA), či zda by musel nejdříve někde v okolí začít hledat fungující telefon. Mentalistický diskurz naproti tomu dokáže velice jednoduše vysvětlit všechny tyto možnosti tím, že daného člověka prostě popíše jako někoho, kdo právě viděl dopravní nehodu, kdo ví, že lidé jsou v ohrožení života a že je potřeba jim zajistit pomoc. Vyjádřeno slovy *Stanislava Komárka* (2000, s. 8), „ač lze zajisté duševní procesy chápat jako projev fyziologických, zejména neurofyziologických daností, je přesto smysluplnější je popisovat jako globálně-autonomní celek a z celistvého pohledu „zvenku“, nežli je rozmotávat zevnitř, asi jako je smysluplnější popisovat druhou světovou válku termíny dějepisy a nikoli jako sumaci pohybu jednotlivých Němců mezi lety 1939 a 1945.“ Proto by podle Pylyshyna (1986) každá vědecká teorie mysli, činící si nárok na vysvětlení chování svědka dopravní nehody, měla být převoditelná na tuto abstraktnější (mentalistickou) úroveň popisu lidského chování. A nejpřímočařejším způsobem, jak toho dosáhnout, je Fodorův reprezentacionalismus, tedy předpoklad, že mozek svědka dopravní nehody skutečně obsahuje nějaké vnitřní fyzikální stavy (neurální symboly či kódy), které reprezentují událost dopravní nehody, a že



"Anthropologists! Anthropologists!"

(výpočetní) přechody mezi jednotlivými vnitřními fyzikálními stavy mozku povedou k dalším vnitřním stavům (tj. k dalším neurálním symbolům či kódům), které by bylo možné interpretovat jako „zavolat na pohotovost“ nebo „najít fungující telefon“. Tímto se již dostáváme k samotné výpočetní teorii mysli, která představuje zatím nejúspěšnější pokus o propojení lidské racionality s kauzálním světem fyzikálních systémů.

Obrázek 3: Lidé dělají určité věci proto, že jsou o něčem přesvědčeni, a proto, že něco chtějí, tedy že mají určité cíle a přání; bez tohoto předpokladu nelze vysvětlit ani to nejjednodušší chování.

1.2 Mysl jako výpočet

V kontrastu s behaviorismem je kognitivní revoluce v americké psychologii na přelomu 50. a 60. let 20. století často pojímána jako znovuoživení mentalistického paradigmatu. Kognitivní revoluce však nebyla pouze návratem k mentalistickému diskurzu, ale především zásadní změnou významu jeho hlavních konceptů, které byly nově redefinovány v rámci tzv. **výpočetní teorie mysli** (*computational theory of mind*), která chápe lidskou mysl jako specifický druh výpočtu realizovaného určitým specifickým typem výpočetního zařízení, kterým je lidský mozek. Tímto tato teorie také odpovídá na otázku, jakým způsobem je možné mechanicky implementovat lidskou racionalitu, resp. jak propojit kauzální svět fyzikálních systémů s jedním ze základních rysů mentálního života, kterým jsou rozumem se řídící posloupnosti mentálních stavů a vzorců chování.

V průběhu dějin psychologického myšlení se objevilo hned několik pokusů přirovnávat lidskou mysl k určitému technologickému zařízení. Takto byla lidská mysl ve své době přirovnávána k hodinovému či parnímu stroji, k telegrafu nebo k telefonní ústředně. Zásadní rozdíl mezi těmito přirovnáními a výpočetní teorií mysli (dále VTM), však spočívá v tom, že v případě VTM se nejedná pouze o nějaké přirovnání či metaforu, ale o konstatování, že lidská mysl je formou výpočtu (v širším smyslu zpracovávání informací, nikoli v užším smyslu numerických výpočtů), resp. počítač, který tento výpočet provádí. K tomuto ztotožnění mysli s počítačem došlo v několika krocích, jejichž počátky můžeme vysledovat hluboko do minulosti. Prvním a nejdůležitějším krokem bylo uvědomění si možností tzv. **formálních systémů**, které, stejně jako člověk, vykazují schopnost generovat „vzorce chování“, které se řídí zákony lidské racionality.

1.2.1 Formální systémy

Pro formální systém jsou charakteristické tři následující vlastnosti (Haugeland, 1997): 1) Jedná se o systém manipulující se znaky (*token-manipulation systems*), tzn. že tento systém je možné plně definovat a) výčtem sady formálních (významu-prostých, nic nereprezentujících) znaků (tj. jakýchkoli objektů jednoznačně identifikovatelných čistě na základě jejich formálních charakteristik jako je tvar, velikost apod.), b) určením jedné nebo více startovních pozic (tj. počátečního uspořádání formálních znaků) a nakonec c) specifikováním souboru formálních pravidel určujících povolené způsoby transformace a manipulace různých konfigurací formálních znaků. 2) Formální systém má digitální povahu,

tzn. že navzdory nepřebornému množství drobných odchylek a variací lze vždy přesně a spolehlivě identifikovat, ve kterém konkrétním stavu - z omezené (povolené) množiny stavů - se nachází jednotlivé prvky systému, které jsou z hlediska jeho fungování důležité.¹

3) Fungování formálního systému je nezávislé na konkrétním médiu, ve kterém je implementován, tzn. že na formálním systému je důležitá pouze jeho forma - jednoznačně rozlišitelné znaky a soubor povolených manipulací a transformací; konkrétní médium - dokud umožňuje implementaci struktury formálního systému - nehraje žádnou zásadní roli.

Klasickým příkladem takového formálního systému, který splňuje všechny tři výše uvedený charakteristiky, jsou šachy, které jsou plně definovány výčtem jednotlivých hracích znaků-figurek, specifikováním jediné možné startovní pozice a výčtem pravidel určujících, jak se mohou jednotlivé figurky po šachovnici pohybovat. Jednotlivé tahy na šachovnici je přitom možné vždy přesně a spolehlivě identifikovat - milimetrové odlišnosti v umístění figurek na šachovnici nevedou ke změnám, které by byly fatální z hlediska správného průběhu celé hry. A nakonec, šachy není potřeba hrát jen s dřevěnou šachovnicí a s dřevěnými figurkami. Šachy je teoreticky možné hrát i s pomocí mezigalaktických korábů a souboru 64 galaxií. Důležité jsou pouze jednoznačně rozlišitelné hrací znaky a soubor povolených tahů, které jsou na těchto znacích definovány.

1.2.1.1 Interpretované symbolické systémy

Typem takového formálního systému jsou i tzv. *interpretované formální systémy*. Zatímco hrací znaky (figurky) šachu nemají žádný konkrétní význam, v interpretovaných formálních systémech znaky svůj význam mají, tzn. že k něčemu odkazují, něco označují nebo reprezentují. Takové znaky se nazývají *symbols* či *symbolické reprezentace*. Formální pravidla pro manipulaci a transformaci symbolických reprezentací (syntax) a významy symbolických reprezentací (sémantika) přitom spolu v rámci interpretovaného formálního systému úzce souvisejí. Tato souvislost je dobře patrná na jednoduchém formálním systému, který ve svém článku jako ilustraci uvádí kognitivní vědec *John Haugeland* (1997, s. 16-17). Jedná se o matematický systém, který zachycuje zákonitosti platící v abstraktním světě čísel a jednoduchých matematických operací jako je sčítání, odčítání, dělení a násobení; skládá se z prvních patnácti písmen anglické abecedy (A-O), z řady několika možných počátečních konfigurací písmen a ze souboru povolených „tahů“, tj. pravidel specifikujících povolené

¹ Základní charakteristikou digitálních systémů, je to, že „její základní elementy jsou buď v jednom stavu, nebo ve stavu jiném, ale nikdy nemohou být někde napůl“ (Dawkins, 2002, s. 121).

způsoby manipulace a transformace jednotlivých písmenných konfigurací. Několik z těchto počátečních konfigurací a povolených „tahů“ je uvedeno v tabulce 1.

Počáteční konfigurace	Povolený "tah"
OEO A	N
NIBM A	G
HCHCH A	KON
KEKDOF A	F
MMCN A	JJ
OODF A	OO
IDL A	M
NBN A	O

Tabulka 1: Několik z povolených herních „tahů“ písmenného matematického systému.(převzato z Haugeland, 1997, s. 16)

To, jestli se skutečně jedná o matematický systém, kriticky závisí na číselných významech jednotlivých písmen. Tato závislost se ozřejmí, když se na tento formální systém podíváme optikou obvyklé symbolické notace, kterou v našem kulturně-civilizačním okruhu používáme ke kódování matematických objektů. V tabulce 2 jsou uvedeny dva možné překlady písmenných symbolů a odpovídající překlady posloupností písmen. Z nich je zřejmé, že transformační pravidla písmenného formálního systému odrážejí zákonitosti světa čísel a základních matematických funkcí pouze v tom případě, kdy jsou písmenům přiřazeny ty správné číselné významy, zde reprezentované indoarabskou číselnou notací.

A=1	I=9	=5=1	/
B=2	J=0	/92x1	7
C=3	K=+	838381	+=/
D=4	L=-	+5+4=61	6
E=5	M=x	xx3/1	00
F=6	N=/	==461	==
G=7	O=	94-1	x
H=8		/2/1	=
A==	I=7	1+1=	2
B=/	J=6	27/3=	9
C=x	K=5	8x8x8=	512
D=-	L=4	5+5-10=	0
E=+	M=3	33x2=	66
F=0	N=2	11-0=	11
G=9	O=1	7-4=	3
H=8		2/2=	1

Tabulka 2: Dva možné překlady písmenných symbolů a odpovídající překlady posloupností písmen uvedených v tabulce 1. (převzato z Haugeland, 1997, s. 17)

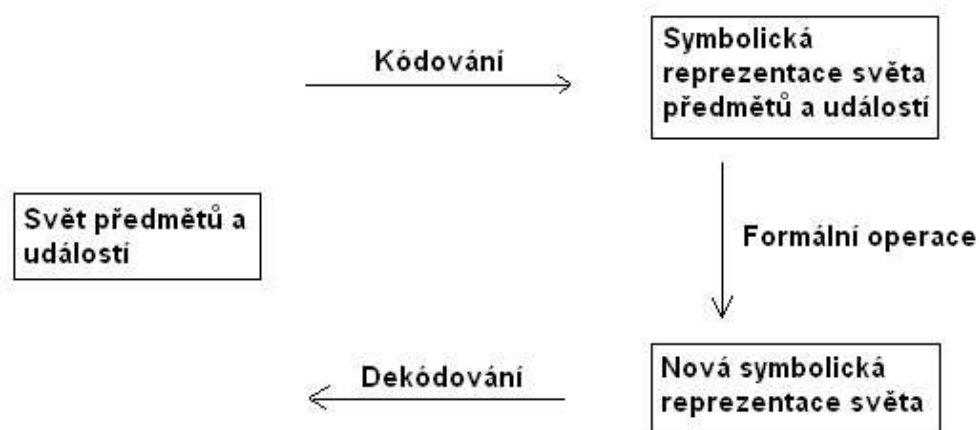
Základním předpokladem interpretovaných formálních systémů je tedy existence symbolických reprezentací, které označují předměty a události v reprezentovaném světě. Tyto symbolické reprezentace na jedné straně zachycují (z hlediska účelu konstrukce daného reprezentačního systému) důležité aspekty reprezentovaného světa a na straně druhé abstrahují od jeho irelevantních detailů. Taková symbolická reprezentace je potom abstraktní a zjednodušenou verzí reprezentovaného světa. Na straně druhé symbolické reprezentace - díky tomu, že mají většinou podobu nějakého konkrétního fyzikálního objektu (například na papír či jiný materiál udělaného klikyháku) - umožňují dát abstraktním objektům a vztahům mezi nimi konkrétní podobu, která je snadno přístupná lidskému sensorickému aparátu. Tato duální povaha symbolických reprezentací je dobře patrná v případě číselné notace, kde jednotlivé znaky na jedné straně abstrahují od konkrétních detailů předmětů, jejichž množství nás zajímá, a na straně druhé dávají konkrétní podobu abstraktním matematickým objektům (číslicům nebo různým matematickým funkcím).



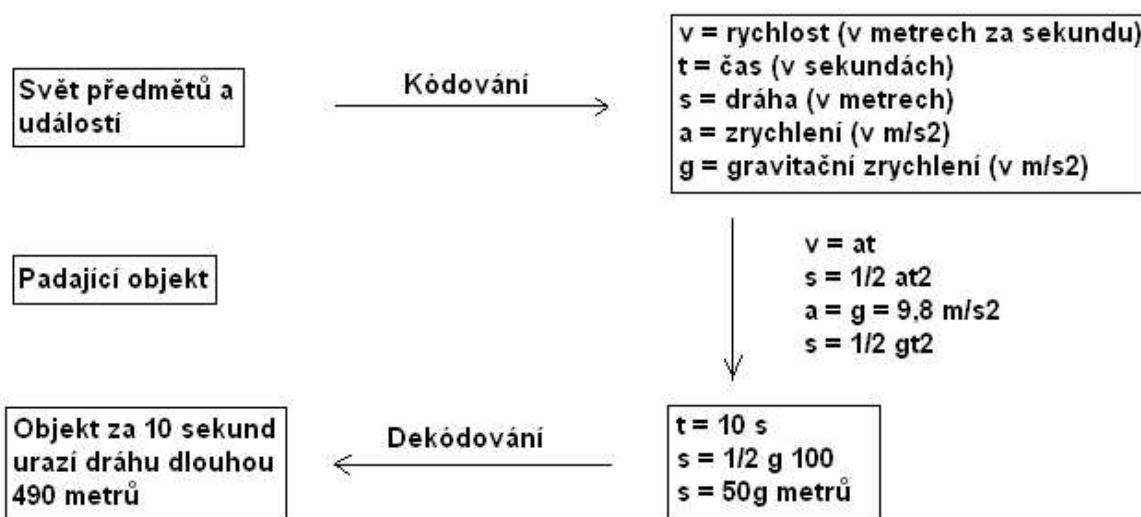
Obrázek 5: *Reprezentační systém. Reprezentovaný svět předmětů a událostí, který v tomto konkrétním případě sestává z šesti lidí, kteří se od sebe v mnoha aspektech liší (jsou například různého věku, pohlaví, jsou různě vysokí nebo mají odlišnou barvu očí), se nachází v horní části obrázku. Svět symbolických reprezentací v podobě jednoduchých značek je pak v dolní části obrázku. Reprezentující svět je zjednodušenou verzí světa reprezentovaného, která abstrahuje od všeho, co není důležité pro účel vytvoření daného reprezentačního systému - v tomto případě jednotlivé značky reprezentují vždy jednu osobu, všechny ostatní aspekty reprezentovaného světa (např. výše zmiňovaný věk, pohlaví, výška či barva očí jednotlivých osob) jsou ignorovány.*

Důležitou vlastností těchto symbolických reprezentací je to, že je možné je uspořádat do (formálního) systému pomocí pravidel, která čistě na základě formálních vlastností symbolických reprezentací (např. na základě jejich tvaru, váhy, barvy apod.) specifikují způsob (formální operace), jakým je možné jednotlivými symboly manipulovat a transformovat je na jiné symboly, tak aby se tyto chovaly způsobem, který odpovídá chování světa, který reprezentují: **formální systém = symboly + pravidla pro manipulaci a transformaci symbolů**

Základní myšlenka interpretovaných formálních systémů tedy spočívá v tom, že zde máme na jedné straně svět předmětů a událostí a na straně druhé určitý způsob systematického kódování těchto předmětů a událostí v podobě různých symbolických reprezentací. S těmito symboly se potom manipuluje způsobem, který odráží zákonitosti platící ve světě předmětů a událostí, který symboly reprezentují. Klíčovou vlastností formálních systémů je to, že manipulace se symboly má formální povahu, tzn. že rozhodnutí o tom, s jakým symbolem a jak manipulovat, jsou činěna čistě na základě odkazu k formě (*syntaxi*) symbolů; významem (*sémantikou*) symbolů, tedy tím, co symboly reprezentují, se není potřeba během manipulací se symboly vůbec zabývat. Výsledkem manipulace se symboly jsou pak další symbolické výrazy a reprezentace, které můžeme nakonec interpretovat jako označující něco ve světě předmětů a událostí. Tato schopnost zachovávat smysluplné sémantické vlastnosti (interpretace) symbolů cestou spoléhání se čistě na jejich formální (syntaktické) vlastnosti činí z formálních systémů významnou *kognitivní technologii* (Clark, 2003; Fauconnier & Turner, 2003; Norman, 1993; Donald, 1991), která značně rozšiřuje dosah lidských kognitivních schopností v čase i prostoru, neboť sestavením toho správného formálního systému (tzn. vytvořením efektivních symbolických reprezentací, které zachycují relevantní předměty a události zájmové domény, a definicí syntaktických pravidel pro manipulaci a transformaci symbolů, které věrně odrážejí zákonitosti platící v zájmové doméně) si člověk může zjednat přístup k předmětům a událostem, které nemá možnost pozorovat přímo, například proto, že jsou příliš daleko, nebo proto, že ještě vůbec nenastaly (viz obrázek 7). Pro nezasvěceného pozorovatele se celý proces nemusí příliš lišit od magického rituálu, kdy se šaman pokouší věštit budoucí události z uspořádání kostí obětního zvířete, které kolem sebe v náboženském tranzu nejdříve rozhodil - i jeho činnost je založena na předpokladu, že existuje určitá korespondence mezi kostmi a světem reálných předmětů a událostí a že určitá nenáhodná („magická“) forma manipulace s těmito kostmi může odhalit něco smysluplného o okolním světě.



Obrázek 6: Formální systém se skládá ze symbolů, které systematicky kódují předměty a události reprezentovaného světa, a z pravidel, která (čistě na základě formy/podoby symbolických reprezentací) specifikují způsob, jakým je možné jednotlivými symboly manipulovat a jak je možné je transformovat na jiné symboly (tzv. formální operace). Tato pravidla přitom odrážejí způsob, jakým se chová reprezentovaný svět, který je kódován systémem symbolických reprezentací. Zajištění této korespondence člověku umožňuje, aby získal přístup k důležitým informacím o světě reálných předmětů a událostí pouze na základě formálních operací prováděných se symboly reprezentujícími tento svět reálných předmětů a událostí (viz obrázek 7). (s úpravami převzato z <http://hci.ucsd.edu/102a/05-lectures/W3a.pdf>)



Obrázek 7: Příklad použití formálního systému. Při zajištění korespondence mezi symboly a pravidly pro jejich manipulaci a transformaci na straně jedné a reprezentovaným světem předmětů a událostí na straně druhé je možné modelovat chování reálného světa. Zde uvedený model člověku umožňuje činit zajímavé a mnohdy i užitečné předpovědi ohledně chování mnoha objektů v dosahu gravitačního pole planety Země. (s úpravami převzato z <http://hci.ucsd.edu/102a/05-lectures/W3a.pdf>)

Tato souvislost mezi formálními pravidly pro manipulaci se symbolickými reprezentacemi a významy symbolických reprezentací sehrála klíčovou roli ve vývoji výpočetní teorie mysli, neboť se ukázalo, že s pomocí interpretovaných formálních systémů je možné úspěšně modelovat zákonitosti platící v mentálním světě významů a logických inferencí prostřednictvím čistě syntaktických (ne-sémantických) nástrojů formálních systémů.

1.2.1.2 Historický a kognitivně antropologický kontext formálních systémů

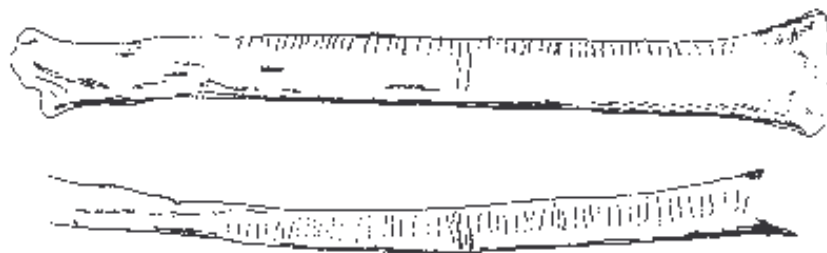
Přestože k explicitnímu teoretickému porozumění formálním systémům došlo až na počátku 20. století díky průkopnickým pracím v oblasti matematiky a logiky - spojeným se jmény jako Bertrand Russel, Alfred Whitehead, David Hilbert, Kurt Gödel, Alan Turing či Alonzo Church –, samotná znalost a používání formálních systémů jsou mnohem staršího data. Jedněmi z prvních a nejdůležitějších formálních systémů byly aritmetické systémy, které jsou známy již z doby čtvrtého tisíciletí před naším letopočtem, kdy je chrámoví a palácoví byrokraté využívali především k vedení účtů o množství zemědělských produktů odevzdaných poddanými na daních; také v současné době je aritmetický systém prvním formálním systémem, se kterým se člověk seznamuje, často ještě před nástupem na základní školu.



Obrázek 4: Početní systém z Mezopotámie z doby okolo roku 3300 př. n. l. Hliněná schránka, na jejímž povrchu jsou vyryty jednoduché značky, které reprezentují obsah schránky, ve které byly umístěny objekty symbolizující množství různých druhů hospodářských plodin a zvířat.

Součástí takového aritmetického systému je vždy určitá číselná notace - sada symbolů (číslic) reprezentujících jednotlivá čísla. Další nezbytnou součástí aritmetického systému je soubor pravidel pro manipulaci a transformaci číselných symbolů odrážejících zákonitosti přechodů mezi čísly v závislosti na použité aritmetické operaci (sčítání, odčítání, násobení a dělení). Vyzbrojen alespoň základní znalostí takového aritmetického systému může pak i žák první třídy směle přistoupit k řešení matematických úloh typu „*Od maminky jsi dostal pět bonbónů a dva z nich jsi dal svému sourozenci. Kolik ti zbylo bonbónů?*“ tím způsobem, že nejdříve přeloží úlohu do symbolické notace $5 - 3$, poté na tento symbolický zápis aplikuje naučená transformační pravidla používaná při odečítání, která symbolický zápis přetransformují do podoby $5 - 3 = 2$, a nakonec číslici 2 přeloží do přirozené „řeči bonbónů“, tak aby mohl odpovědět, že mu zbyly právě dva bonbóny.

Každý typ číselné notace používané k symbolickému kódování kvantitativních charakteristik vnějšího světa umožňuje různé druhy manipulací a transformací, které jsou v různé míře vhodné k implementaci základních aritmetických operací. Nejjednodušší a nejstarší (viz obrázek 8) známou číselnou symbolickou reprezentací jsou jednoduché značky v podobě svislých čárek označujících vždy jednu položku. Jedná se o tzv. *aditivní číselnou notaci* (Norman, 1993) - když chce člověk zvýšit hodnotu předchozího číselného symbolu, jednoduše přidá k tomuto symbolu další symbol; tímto způsobem se například ze symbolu pro číslo 3 (III) stane symbol pro číslo 4 (IIII). Výhodou tohoto způsobu kódování je to, že se relativně snadno vytváří, díky čemuž jím lze rychle a pohodlně zaznamenávat počet jednotlivých položek. Z tohoto typu číselné notace lze také relativně snadno vyčíst konkrétní číselnou hodnotu, kterou daný symbol reprezentuje – k usnadnění tohoto dekodování se často jednotlivé čárky seskupují po pěti (IIII). Vedle těchto předností je další výhodou to, že velikost (délka) tohoto typu číselné reprezentace je proporční vzhledem k velikosti čísla, které označuje, takže číselný symbol pak může fungovat de facto jako jistý druh grafu, z něhož lze – v případě, že takto reprezentujeme několik čísel najednou - přímo vyčíst relativní velikost reprezentovaných čísel. Právě tyto vlastnosti jsou důvodem, proč se tento typ číselné notace používá až do dnešních dnů. Jeho zásadní nevýhodou je však to, že nepodporuje ty druhy pravidel pro manipulaci a transformaci číselných symbolů, které by dokázaly efektivně implementovat složitější aritmetické výpočetní operace s většími čísly.



Obrázek 8: Věstonická vrubovka. V roce 1936 u obce Dolní Věstonice na archeologickém nalezišti z období mladšího paleolitu (zhruba před 25 000 lety) našel archeolog Karel Absolon osmnáct cm dlouhou lýtkovou kost mladého vlka, na které bylo vyryto 55 zářezů - od dvou delších, ležících uvnitř řady, jich bylo na jednu stranu 25 a na druhou 30. Zářezy podle všeho představují číselné pojmy, násobky pěti (pěti prstů na ruce?), jednou pět x pět, podruhé šest x pět. Název nálezu (vrubovka) je odvozen od názvu holí či latěk, které svého času sloužily jako pomůcky k účtování (také „rabuše“ a „rováš“). Dělaly se do nich zářezy, vruby, podle počtu dodaného zboží, pracovních jednotek apod. Po dodání a zaplacení zboží se vrubová plocha opět vyhladila. Existovaly také vrubovky dvojité. Když bylo třeba potvrdit nějakou obchodní transakci, části vrubovky obou zúčastněných stran se přiložily k sobě a vrub se udělal najednou. Při účtování pak musely vruby souhlasit, jeden přiléhal k druhému. Později se „vrubovka“ říkalo také tabulím v hospodách, jak to dokládají slovní obraty typu „udělat to na svůj vrub“ (na svůj účet, na svou odpovědnost), „má u mne vroubek“ (nevyrovnaný dluh). (Folta, Loucká, Klíma, 1997)

V tomto ohledu jsou mnohem efektivnější číselnou notací římské nebo v současné době zcela převládající indoarabské číslice. Indoarabská číselná notace je příkladem tzv. *substituční reprezentace* (Norman, 1993) - při zvyšování hodnoty předchozího symbolu musí člověk tento symbol nejdříve vymazat a potom nahradit jiným symbolem – takto se například při zvyšování hodnoty číselného symbolu z 1 stane \top 2, z \top 2 se stane $\top\top$ 3 atd. Další důležitou vlastností indoarabské číselné notace je to, že význam, resp. hodnota každého číselného symbolu závisí na jeho pozici a umístění, což je také důvod, proč indoarabský číselný systém potřebuje nulu - aby 3 mohla znamenat „300“, musí se nacházet na třetí pozici zprava. Římský číselný systém, který v sobě obsahuje prvky jak substituční tak i aditivní reprezentace a je de facto pouze mírnou modifikací jednoduchého „čárkování“, naproti tomu nulu nepotřebuje, neboť jeho jednotlivé číselné symboly reprezentují vždy stejnou číselnou hodnotu bez ohledu na jejich pozici – alespoň v případě původní podoby římské číselné notace, ve které se číslo 4 zaznamenávalo jako IIII a číslo 9 jako VIIII a ve které nezáleželo na pořadí, v jakém byly jednotlivé číselné symboly uspořádány, takže MMVI mělo stejnou hodnotu jako VMIM. Oba systémy mají – z hlediska výpočetních operací, které podporují – své silné i slabé stránky. V odborné literatuře se často zmiňuje aritmetická operace násobení jako doklad výhod, které nám skýtá užívání indoarabského číselného systému ve srovnání s

římskou notací – vynásobení dvou čísel, které jsou zapsány pomocí římských číslic, např. MMCLIII a CLXXXII, je sice možné, ale velice obtížné ve srovnání s násobením stejných čísel zapsaných v indoarabské notaci (2153 a 182). Již méně se ví, že operace sčítání je naopak mnohem jednodušší s římskými číslicemi: Když chce člověk sčítat, ať už s pomocí římských nebo indoarabských číslic, musí se nejdříve naučit jednotlivé číselné symboly a pravidla pro jejich manipulaci a transformaci. V případě indoarabské číselné notace to znamená, že

- člověk se musí naučit deset arbitrární symbolů pro deset čísel (0 až 9);
- dále si musí osvojit pravidla týkající se způsobu kódování jednotlivých číselných hodnot prostřednictvím zápisu číselných symbolů na různé pozice, tak aby věděl, že například 93 se rovná devíti desítkám a třem jednotkám;
- dále se musí naučit nazpaměť tabulku součtů všech 45 možných párů čísel (deset číslic může vytvořit 100 párových kombinací, avšak vzhledem k platnosti komutativního zákona, kdy $a + b = b + a$, a vzhledem k tomu, že $a + 0 = a$, je nutné se naučit „pouze“ 45 různých kombinací);
- a nakonec se člověk musí naučit, co dělat, když dochází k přechodu z jednoho početního řádu do druhého.

Naučit se tohle všechno je dost namáhavé a zabere to hodně času. Naproti tomu v případě římské číselné notace člověku stačí naučit se pouze

- sedmi arbitrárními znakům pro číselné hodnoty od 1 do 1000 (I, V, X, L, C, D, M);
- jednoduché kombinaci jednotlivých symbolů a
- jejich přeuspořádání tak, aby stejné symboly stály vedle sebe, a nakonec
- aplikaci několika málo jednoduchých pravidel, která komplexní číselnou reprezentaci zjednoduší (např. IIIII = V, XXXXX = L atd.).

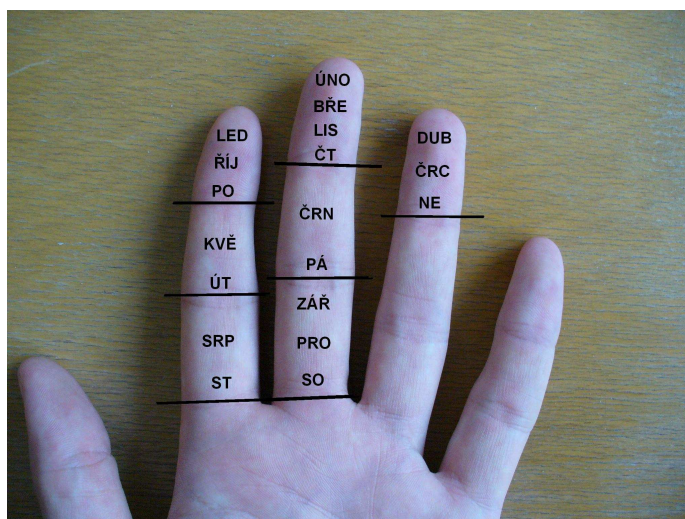
Tato pravidla pro manipulaci a transformaci číselných symbolů spojená s římskou číselnou notací si lze osvojit mnohem snáze a rychleji než ta spojená s indoarabským číselným systémem. Například při sčítání dvou čísel reprezentovaných jako MMCLIII a CLXXXII stačí symboly nejdříve jednoduše zkombinovat (MMCLIIICLXXXII), potom přeuspořádat (MMCCLLXXXIIII), nakonec celý výraz zjednodušit (MMCCCXXXV) a přečíst konečný výsledek (dvatisícetřicetpět). V tomto ohledu to měli římské děti mnohem jednodušší než ty dnešní – ovšem jen do chvíle než přišlo na násobení (nebo dělení). Zde je výhodnější použít indoarabské číselné notace, která - na rozdíl od té římské - činí na první pohled zřejmou a tím snadno dostupnou informaci o početním řádu jednotlivých čísel (jednotky,

desítky, stovky...), která slouží jako jeden z podstatných vstupů do výpočetní procedury dlouhého násobení – budeme-li chtít mezi sebou vynásobit například čísla 2153 a 182, potom stačí s pomocí znalosti malé násobilky jednoduše vynásobit každou číslici v prvním čísle (2, 1, 5, 3) s každou číslicí v druhém čísle (1, 8, 2), vypsat výsledky na ta správná místa, která kódují početní řády jednotlivých čísel, a nakonec takto zapsané dílčí výsledky sečíst (Norman, 1993).

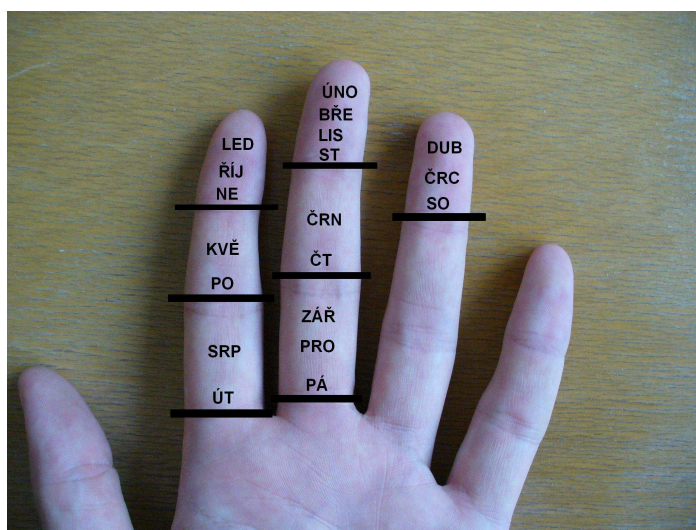
Vzhledem ke komplexnosti pravidel pro manipulaci a transformaci symbolů v rámci formálních systémů a s tím spojenými nároky na kapacitu pracovní paměti člověka, který s formálním systémem pracuje, zde vždy existoval tlak na „krystalizaci“ těchto systémů v podobě nejrůznějších kognitivních artefaktů (Hutchins, 1995). Kognitivní antropolog *Edwin Hutchins* (2004) v této souvislosti hovoří o tzv. *materiálních kotvách* (*material anchors*) – předmětech, na jejichž fyzikální strukturu je namapována abstraktní struktura určitého konceptuálního modelu se všemi jeho jednotlivými prvky a vztahy mezi nimi. Taková materiální kotva je potom de facto ztělesněním či hmotnou „inkarnací“ zákonitostí a omezení, která doposud existovala pouze na abstraktní rovině v podobě nějakého mentálního modelu - zde ve smyslu obecného vnitřního modelu vnějšího světa, v němž jsou zachyceny podstatné prvky světa a vztahy mezi nimi, bez ohledu na konkrétní podobu reprezentačního kódu, tedy bez ohledu na to, zda jsou dané prvky a zákonitosti reprezentované prostřednictvím propozičního kódu, analogového kódu, nebo kombinací obého. Díky ukotvení modelu ve struktuře hmotného artefaktu dochází ke zvýšení stability jeho reprezentace, která je klíčová pro přesnost a efektivitu lidského usuzování (Sternberg, 2002). Jednotlivé prvky a zákonitosti takto ukotveného modelu jsou potom stabilně reprezentované i během komplexních procesů manipulace a transformace, kterým je model v průběhu řešení daného kognitivního úkolu podrobován.

Ilustrací efektu ukotvení abstraktní struktury modelu na fyzikální strukturu hmotného předmětu je jednoduchý ruční kalendář (Hutchins, 2004), který usnadňuje jistý druh výpočtů nutných k určení dne (Po-Ne), který připadá na konkrétní datum v jakémkoli týdnu jakéhokoli měsíce daného roku. Kdokoli, kdo by chtěl tento typ problému vyřešit, musí určitým specifickým způsobem zkombinovat a zkoordinovat prvky a zákonitosti obsažené v týdenním cyklu sedmi dní a ve dvanácti měsíčních cyklech 28 (resp. 29, v případě, že se jedná o přestupný rok), 30 nebo 31 dní. Vzhledem ke komplexnosti výsledného schématu je potřeba ho ukotvit k nějaké pevné fyzikální struktuře. V případě ručního kalendáře je touto strukturou články prvních tří prstů na levé ruce (počítáno od ukazováčku). Z devíti článků, které jsou na těchto třech prstech, je využíváno pouze sedm. Do každého ze sedmi článků jsou podle

určitého pravidla umístěny názvy několika měsíců (viz obrázek 9a, b). Tato posloupnost názvů měsíců v jednotlivých člancích je pevná a neměnná, tzn. že pořadí měsíců na prstech je stabilní a v průběhu let se nijak nemění. Protože je tato posloupnost poměrně složitá a obtížná na zapamatování v podobě nějaké vizuální představy, lidé, kteří ruční kalendář používají, se většinou spoléhají na „svalovou paměť“, kdy se posloupnost měsíců učí jako posloupnost pohybů palce levé ruky přes jednotlivé články. Názvy dní jsou v jednotlivých člancích umístěny vždy po jednom v rámci snadno zapamatovatelné posloupnosti „odshora-dolů“ a „zleva-doprava“. Na rozdíl od posloupnosti měsíců není tato posloupnost v průběhu let neměnná - každý následující rok se celá struktura posune o jednu pozici, takže jestliže v konfiguraci pro rok 2002 je na prvním článku umístěno pondělí (obrázek 9a), v roce 2003 to bude úterý, v přestupném roce 2004 to ale bude čtvrtek (+ 2 dny), v roce 2005 pátek, v roce 2007 neděle (obrázek 9b) atd. Výsledný průnik těchto dvou struktur dává vzniknout nové struktuře, která zachycuje vztah mezi týdenním cyklem sedmi dní a měnící se délkou jednotlivých měsíčních cyklů. Díky tomu, že týden má vždy sedm dní, všechny dny v jakémkoli měsíci, jejichž datum se rovná 0 modulo 7 (tj. 7., 14., 21. a 28. den měsíce), připadají v rámci týdne na stejný den; například v roce 2007 připadají 7., 14., 21. a 28. den měsíce září vždy na pátek. Samotný výpočet s pomocí ručního kalendáře pak probíhá následujícím způsobem: Řekněme, že chceme zjistit, jaký den bude připadat na 24. prosince roku 2007. Člověku stačí se jednoduše podívat na svou levou ruku, přitom si představovat názvy dní a měsíců umístěných v jednotlivých člancích, a najít ten článek, ve kterém je obsažen název měsíce, který ho zajímá, v tomto konkrétním případě to bude prosinec. Název dne v tomto článku je „pátek“, tzn. že v prosinci roku 2007 připadají všechny pátky na dny s datem, které se rovná 0 modulo 7 (7., 14., 21. a 28. prosinec roku 2007). Nyní tedy víme, že 21. prosince roku 2007 bude pátek. S touto informací je již velice snadné určit, že 24. prosince bude pondělí – 24. prosinec je o 3 dny dále než 21. prosinec, takže stačí dopočítat 3 dny a zjistit název dne, který je obsažen v příslušném článku. Tímto způsobem je možné rychle a přesně určit den připadající na jakékoli datum během kalendářního roku – stačí určit nejbližší známý den se známým datumem (0 modulo 7, tj. 7., 14., 21. a 28. den měsíce) a dopočítat, nebo odpočítat zbývající, resp. přebývající dny. Bez podpůrné struktury sedmi článků na prstech levé ruky by však bylo velice obtížné udržet v pracovní paměti stabilně reprezentované komplexní vztahy mezi týdenním cyklem a měsíčními cykly a provádět výše uvedené manipulace a transformace spojené s řešením dané výpočetní úlohy.



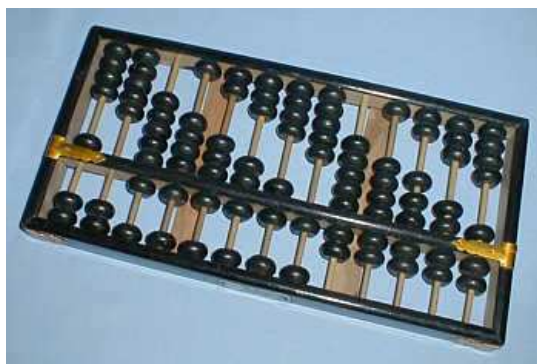
Obrázek 9a: „Ruční kalendář“. Konfigurace pro rok 2002.



Obrázek 9b: „Ruční kalendář“. Konfigurace pro rok 2007.

V případě ručního kalendáře sloužila materiální kotva jako podpůrná struktura pro stabilizaci symbolických reprezentací komplexních vztahů mezi týdenními a měsíčními cykly. Jakkoli je materiální kotva užitečnou pomůckou, která snižuje nároky kladené na pracovní paměť, samotnou manipulaci a transformaci symbolických reprezentací musí člověk i nadále vykonávat ve své mysli. Existuje však jistá třída kognitivních artefaktů, které umožňují převést tuto poměrně obtížnou mentální činnost na jednoduchou fyzickou manipulaci a transformaci stavů určité fyzikální struktury, která implementuje prvky a zákonitosti nějakého abstraktního modelu, např. aritmetického systému.

Klasickým příkladem takového kognitivního artefaktu je **abakus**¹, stará, ale dodnes používaná výpočetní pomůcka, která pro lidskou mysl náročnou posloupnost manipulací a transformací abstraktních symbolů aritmetického systému redukuje na posloupnost jednoduchých fyzických manipulací s korálky, jejichž vzorce v různých sloupcích reprezentují různé číselné hodnoty. Aby abakus mohl fungovat jako počítadlo, musí se lidé naučit takovým pravidlům pro transformování a manipulaci těchto vzorců korálků, která zajišťují, že sémantická interpretace dvojic vstupní vzorec korálků-výstupní vzorec korálků odpovídá určité užitečné matematické funkci jako je sčítání nebo odčítání. Na samotných pravidlech přitom není de facto nic matematického - jsou to pouze pravidla pro pohybování korálky. To, co činí tato pravidla užitečnými, je to, že je zajištěna neustálá korespondence mezi těmito formálními pravidly pro pohybování korálky a matematickými objekty jako jsou čísla nebo nějaké matematické funkce (Pylyshyn, 1989). Jinými slovy, díky fyzikální



Obrázek 10: *Abakus redukuje pro lidskou mysl náročnou výpočetní činnost na poměrně jednoduchá formální pravidla fyzické manipulace s korálky, která proměňují jeden vzorec korálků na jiný.*

strukturu počítadla a jednoduchým pravidlům, kterými se člověk řídí při manipulaci a transformaci stavů této fyzikální struktury, se počítadlo chová způsobem, který odráží zákonitosti platící v abstraktní doméně čísel a matematických funkcí jako je sčítání nebo odčítání. (Na přiloženém CD je možné pod názvem „Abakus“ nalézt krátkou videoukázku, na které je zachycen postup při sčítání několika třiciferných čísel s pomocí abaku.)

Jiným takovým dobře známým kognitivním artefaktem je **logaritmické pravítko**, které představuje fyzikální implementaci matematických zákonitostí, které umožňují zjednodušit aritmetické operace dělení a násobení prostřednictvím jejich transformace na aritmetické operace odčítání a sčítání logaritmů čísel, která mezi sebou chceme vydělit nebo vynásobit:

¹ Abakus je také znám pod názvem *soroban* (Japonsko), *suan-pan* (Čína), *sčot* (Rusko), *coulba* (Turecko) nebo *choreb* (Arménie) (Seife, 2005).

$$\log(ab) = \log a + \log b$$

$$\log(a/b) = \log a - \log b$$

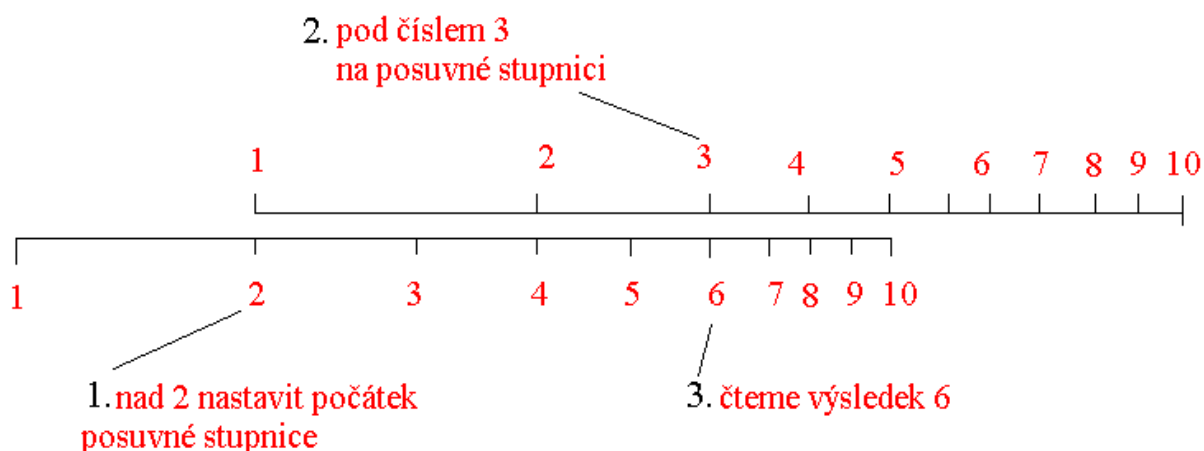
Fyzikální struktura logaritmického pravítka (které se většinou skládá ze tří částí - základní část neboli těleso, posuvná část neboli šoupátko a posuvný průhledný rámeček neboli běžec s jednou nebo více vlasovými ryskami zvanými indexy, viz obrázek 11) pak realizuje aritmetické operace dělení a násobení prostřednictvím sčítání a odčítání prostorových délek, které reprezentují logaritmy násobených nebo dělených čísel. Uživatel tak – stejně jako v případě abaku – de facto nedělá nic matematického, pouze jistým specifickým způsobem fyzicky manipuluje s prostorovými délkami zakódovanými v komponentách logaritmického pravítka; přesto je výstupem celého procesu – po náležité interpretaci – smysluplná odpověď na matematickou otázku, jaký je výsledek násobení nebo dělení dvou čísel. (Pod názvem „Log_pr“ je na příloženém CD umístěna krátká videoukázka, na které je zachycena procedura násobení dvou čísel s pomocí logaritmického pravítka.)



Obrázek 11: Logaritmické pravítko.

Postup při násobení na logaritmickém pravítku

Počítáme 2×3 --- $\log 2 + \log 3 = \log 6$



Obrázek 12: Násobení čísel na logaritmickém pravítku.

Důležitou vlastností kognitivních artefaktů jako je abakus nebo logaritmické pravítko je to, že člověku umožňují provozovat matematiku aniž by přitom člověk musel dělat cokoli matematického v obvyklém slova smyslu. To je umožněno tím, že při používání daného kognitivního artefaktu se člověk stává součástí širšího, **distribuo­vaného kognitivního systému** (Hutchins, 1995), jehož funkce se odlišuje od funkce, kterou v tomto systému hraje jedna jeho dílčí (jakkoli klíčová) část, tedy člověk. Jinými slovy, je třeba rozlišovat mezi tím, jakou kognitivní funkci realizuje daný distribuovaný kognitivní systém jako celek (složený z lidského subjektu a kognitivního artefaktu), a tím, jaký typ kognitivní úlohy stojí před samotným člověkem. Je to vlastně jakási forma dělby (kognitivní) práce, kdy každá jednotlivá součástka distribuovaného kognitivního systému přispívá svým vlastním omezeným způsobem k funkci celého systému.

Při popisu vztahu celkové funkce distribuovaného kognitivního systému a dílčích kognitivních nároků kladených na jeho lidskou komponentu je velice užitečný koncept tří úrovní analýzy kognitivních systémů, který ve své práci o zrakovém vnímání zavedl *David Marr* (1982): První úrovní analýzy každého kognitivního systému by podle Marra měla být tzv. **výpočetní** (také znalostní, sémantická či pojmová) **úroveň analýzy** úkolu nebo funkce, kterou daný kognitivní systém vykonává. Na této úrovni analýzy by mělo být jednoznačně specifikováno, co daný kognitivní systém dělá, jak to dělá a jaké informace k tomu využívá. V informačně-procesní terminologii tato úroveň analýzy představuje specifikaci vstupně-výstupní funkce (tedy co je vstupem do kognitivního systému a jaký je z něj výstup). Druhá, tzv. **reprezentační a algoritmická** (také symbolická či syntaktická) **úroveň analýzy** spočívá ve specifikaci způsobu reprezentace či kódování vstupních a výstupních informací a v popisu algoritmu, který transformuje vstupní informace na výstupní informace. A konečně na třetí, tzv. **implementační** (také fyzické či biologické) **úrovni analýzy** je specifikován konkrétní způsob, jakým jsou dané reprezentace a algoritmy fyzikálně implementovány. Aplikaci této tříúrovňové analýzy lze ilustrovat na jednoduchém kognitivním systému, jakým je abakus: Na výpočetní úrovni analýzy lze abakus popsat jako systém, který provádí aritmetické výpočty, tzn. že na výstupu poskytuje řešení aritmetických problémů (sčítání a odečítání čísel), které jsou mu předloženy na vstupu. Reprezen­tační a algoritmická úroveň analýzy abaku by sestávala z popisu pravidel, kterými se člověk řídí při manipulaci s korálky během sčítání, a odečítání čísel. Implementační úroveň popisu by pak závisela na tom, z jakého materiálu by byl abakus vyroben – mohly by to být dřevěné kuličky navlečené na kovových drátech zapuštěných do dřevěného trámu, tenisáky v pokoji na prkenné podlaze, nebo také systém tvořený tužkou, linkovaným papírem a dobrou gumou. Na výpočetní úrovni analýzy je abakus

totožný se všemi běžnými výpočetními pomůckami jako je logaritmické pravítko nebo kapesní elektronická kalkulačka. Na reprezentační, algoritmické a implementační úrovni analýzy by se však tyto jednoduché kognitivní systémy od sebe již významně lišili, neboť každý z nich využívá při řešení aritmetických problémů odlišnou strategii.

Tento poznatek, že „když dva dělají totéž, nemusí to být vždy totéž“, lze úspěšně využít i při analýze distribuovaného kognitivního systému (sestavajícího z lidského „operátora“ a nějakého kognitivního artefaktu) a při osvětlení transformace problémové domény, ve které se lidská mysl pohybuje, v důsledku zařazení člověka do různých kognitivních funkčních systémů. Jako ilustraci zde můžeme využít Hutchinsův (1995) příklad s počítáním rychlosti lodi v závislosti na čase a vzdálenosti, kterou loď za tento čas urazí: *Lod' urazila 1500 yardů za 3 minuty. Jaká je její rychlost v uzlech/námořních mílích za hodinu?* Vzhledem k tomu, že se jedná o úlohu, kterou je lodní navigátor nucen během plavby velice často řešit, navíc často pod značným časovým tlakem (při plavbě v blízkosti pobřeží se rychlost lodi měří v 1, 2 a 3-minutových intervalech), vždy zde existovala silná poptávka po postupech, které by zaručily rychlé a zároveň přesné a spolehlivé řešení této výpočetní úlohy.

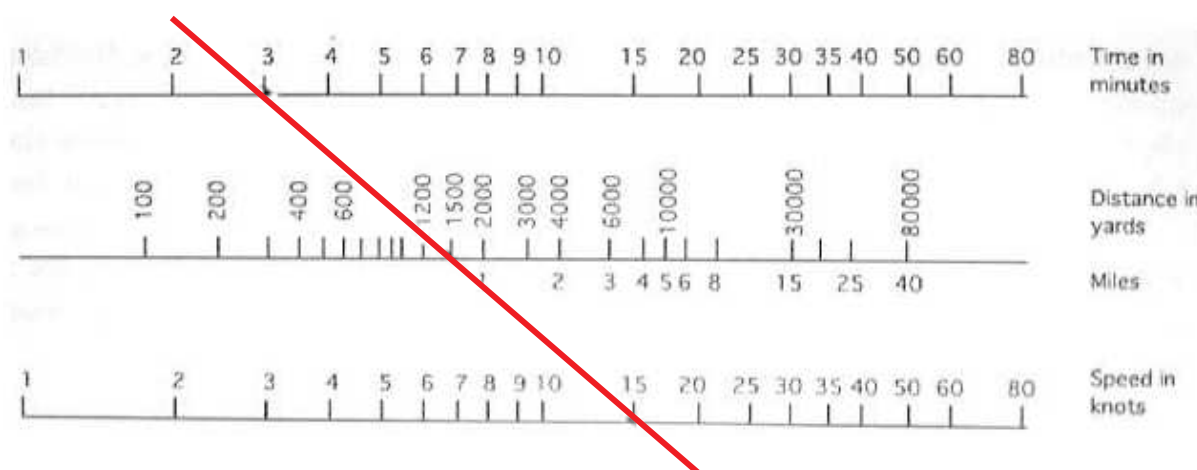
Danou úlohu může navigátor počítat čtyřmi různými způsoby: s pomocí 1) tužky a papíru, 2) kapesní elektronické kalkulačky, 3) tříškálového nomogramu, nebo s pomocí 4) tzv. třiminutového pravidla. S každým z těchto kognitivních artefaktů, které mají buď podobu materiální struktury (1, 2 a 3), nebo mentálního algoritmu či heuristiky (4), navigátor vytváří čtyři různé funkční kognitivní systémy. Na výpočetní úrovni analýzy jsou všechny tyto čtyři funkční systémy identické: všechny jsou schopny přetransformovat vstupní informaci o upluté vzdálenosti (v yardech) a uplynulém čase (v minutách) na výstupní informaci o rychlosti lodi (v uzlech/námořních mílích za hodinu), tzn. že všechny plní stejnou funkci a všechny dávají odpověď na stejnou otázku. Totožné jsou rovněž v tom, že všechny musí nějakým způsobem implementovat zákonitosti týkající se vztahu mezi dráhou, rychlostí a časem ($\text{dráha} = \text{rychlost} \cdot \text{čas}$) a mezi jednotlivými jednotkami času a délky (60 minut = 1 hodina, 1 námořní míle = 2000 yardů). Zásadní rozdíl mezi těmito funkčními systémy však spočívá v mediačních strukturách a procesech, které jednotlivé funkční systémy využívají při transformaci vstupních informací na výstupní údaje o rychlosti lodi; liší se zejména ve způsobu, jakým reprezentují informace, se kterými během výpočtu pracují, a potom v transformačních algoritmech, které aplikují na tyto reprezentace, aby vyprodukovaly požadovanou výstupní informaci. Součástí či důsledkem těchto rozdílů mezi jednotlivými funkčními systémy jsou pak i odlišné kognitivní nároky, které na navigátora daný typ úlohy klade.

V případě počítání s pomocí tužky a papíru navigátor potřebuje znalosti algebry, aritmetiky a již výše uvedené poznatky, že 1 námořní míle = 2000 yardů, 60 minut = 1 hodina a že dráha (s) = rychlost (v) * čas (t). Tyto reprezentační struktury jsou koordinovány následujícím způsobem: Navigátor musí nejdříve použít své znalosti algebry, aby přetransformoval rovnici $s = v * t$ do podoby $v = s/t$, tak aby rychlost mohla být přímo vypočítána z daných hodnot s a t . Potom musí být vzdálenost v yardech převedena na ekvivalentní počet mílí a to za použití znalosti počtu yardů v míli a znalostí aritmetiky; podobně musí být převeden čas s pomocí znalosti počtu minut v hodině a znalostí aritmetiky na ekvivalentní časový údaj v hodinách. Nakonec stačí vydělit upravenou vzdálenost upraveným časem a navigátor získá výsledný údaj o rychlosti lodi v uzlech/námořních mílích za hodinu. Všechny tyto kombinace jednotlivých reprezentačních struktur mohou být provedeny trochu jiným způsobem (v jiném pořadí), ale všechny musí být v tomto funkčním systému obsaženy. To, co je na tomto postupu obtížné, a tedy i více náchylné k chybám, nejsou jednotlivé dílčí výpočty, ale především jejich správná kombinace a seřazení.

Při použití kapesní elektronické počítačky člověk využívá stejné reprezentační struktury jako v předchozím funkčním systému. To, co se změnilo, jsou pouze postupy, kterými navigátor realizuje aritmetické operace násobení a dělení - místo vytváření symbolických struktur na kusu papíru a rozkládání problému na sadu operací prováděných na jednočíselných aritmetických výrazech, navigátor vkládá dané číselné hodnoty a provádí aritmetické operace s těmito číselnými hodnotami prostřednictvím mačkání tlačítek kalkulačky v odpovídajícím pořadí. Oproti předchozímu postupu tak kalkulačka zjednodušuje tu část úkolu, která je jednoduchá; tu obtížnou (správnou koordinaci a posloupnost jednotlivých operací) však navigátorovi nijak neulehčuje.

Při použití tříškálového nomogramu (obrázek 13) se však celá úloha pro navigátora dramaticky zjednoduší. Navigátorovi postačí, když si na odpovídajících škálách najde hodnoty časového údaje (v minutách) a údaje o vzdálenosti (v yardech) a obě tyto hodnoty si na nich označí. Takto označené hodnoty pak propojí pomocí pravítka a tam, kde se bude třetí škála s pravítkem protínat, navigátor najde výslednou hodnotu rychlosti lodi v uzlech/námořních mílích za hodinu. K realizaci celého výpočtu takto navigátor nepotřebuje žádné z výše uvedených znalostí o převodu jednotek, ani znalosti algebry nebo aritmetiky. Tím, že jsou všechny škály v těch správných jednotkách, odpadá problém s jejich převodem do správného formátu; k transformaci $s = v * t$ na $v = s/t$ navigátor nepotřebuje žádné znalosti syntaxe algebraických transformací, neboť vztahy mezi jednotlivými prvky rovnice ($s = v * t$, $v = s/t$ a $t = s/v$) jsou ztělesněny v materiální struktuře nomogramu, takže celé výpočetní

plánování je nahrazeno jednoduchou manipulací hmotným artefaktem. Veškerou náročnou kognitivní práci – v podobě algebraických a aritmetických transformací – za navigátora odvede nomogram, resp. jiná osoba, která v minulosti vtělila své znalosti do struktury nomogramu (tzv. *v čase distribuovaný kognitivní systém*). Díky specifickému způsobu reprezentace a transformace informací nutných k řešení dané úlohy se tak významně změnil profil „kognitivních investic“, které musí navigátor vložit do celého projektu: Ve srovnání s předchozími dvěma funkčními systémy navigátor získá odpověď na svou otázku prostřednictvím nenáročných (protože většinou automaticky a nevědomě probíhajících) kognitivních operací spojených se zrakovým vnímáním, orientací v prostoru a senzomotorickou koordinací. Nomogram tak významně změnil a zjednodušil problémovou doménu, ve které se navigátor musí pohybovat, když chce vědět, jaká je rychlost lodi v námořních mílích za hodinu, když ví, že loď uplula tolik a tolik yardů za tolik a tolik minut.



Obrázek 13: Tříškálový nomogram. (převzato z Hutchins, 1995, s. 148)

Podobně jako nomogram i tříminutové pravidlo zásadním způsobem mění povahu problému, kterému čelí navigátor při počítání rychlosti lodi (v uzlech) na základě známých údajů o uplynulém časovém úseku (v minutách) a vzdálenosti, kterou loď za tento časový úsek urazila (v yardech). Avšak zatímco nomogram představuje hmotný artefakt, tříminutové pravidlo je vnitřním, mentálním artefaktem, který má podobu jednoduchého pravidla či algoritmu, na základě kterého se navigátor dokáže prostřednictvím nijak obtížných kognitivních operací dostat rychle od vstupních informací k požadovanému výstupu. Tříminutové pravidlo využívá skutečnosti, že 3 minuty jsou $1/20$ hodiny a 100 yardů představují $1/20$ námořní míle, takže počet stovek yardů, které loď upluje za 3 minuty, se rovná rychlosti lodi v námořních mílích za hodinu. Jestliže tedy loď za 3 minuty upluje 1500

yardů, potom loď pluje rychlostí 15 námořních mil za hodinu, resp. rychlostí 15 uzlů. Stejně jako při použití nomogramu, ani zde navigátor de facto nic nepočítá – alespoň ne v tradičním slova smyslu. Odpověď na svou otázku jednoduše „vidí“ a to v zásadě okamžitě poté, co zjistí, kolik yardů loď uplula za tři minuty. Navigátorovi stačí představit si obvyklý symbolický zápis upluté vzdálenosti (v yardech), ze kterého jsou odstraněny poslední dvě číslice. To je veškerá duševní práce, kterou musí navigátor - vybavený tříminutovým pravidlem - provést.

Je celkem všeobecně známo, že schopnost člověka vyřešit určitý problém do značné míry závisí na způsobu, jakým je daný problém reprezentován. V této souvislosti se hovoří o tzv. *reprezentačním efektu* (Zhang, Norman, 1994), který spočívá v tom, že různé způsoby reprezentace jedné společné abstraktní struktury problému mohou vést k dramaticky odlišnému kognitivnímu chování, tj. k zapojení odlišných kognitivních procesů a struktur nebo k zapojení stejných kognitivních zdrojů, ale v odlišné míře. Právě tato skutečnost, že různé způsoby reprezentace kladou různé požadavky na kognitivní schopnosti člověka, stojí v pozadí vlivu způsobu reprezentace na míru obtížnosti řešení daného problému. To, do jaké míry je určitý způsob reprezentace efektivní a uživatelsky vstřícný a do jaké míry člověku pomáhá porozumět povaze problému a následně ho tedy i úspěšně vyřešit, závisí na tom, do jaké míry zachycuje důležité aspekty reprezentovaného problému a ignoruje ty aspekty problémové situace, které jsou z hlediska jejího řešení irelevantní. To nejdůležitější hledisko je však míra explicitnosti, resp. implicitnosti zakódování relevantních informací v dané reprezentaci - jak snadno a rychle, tj. s vynaložením jak velké kognitivní námahy v podobě různých mentálních operací, se člověk prostřednictvím takového způsobu reprezentace dostane k informacím, které jsou z hlediska řešení daného problému důležité. Způsob reprezentace informace totiž ovlivňuje „nákladovou strukturu informace“, tzn. že mění míru a povahu „kognitivních investic“, které jsou nutné k získání takové informace - každý způsob reprezentace totiž určitou informaci dává do popředí a jinou upozaduje, takže zatímco některou informaci je velice snadné získat, protože leží jakoby na povrchu, jinou informaci je potřeba doslova „vydestilovat“ prostřednictvím dalších kognitivních operací. Čím snadněji a rychleji lze danou informaci získat, čím je bezprostředněji přístupná, čím méně je potřeba k jejímu získání či k její interpretaci dalších mentálních operací a struktur, tím více je taková informace explicitně reprezentována (Kirsh, 1990). Např. číslice „25“ reprezentuje číslo 25 explicitněji než „5²“, přestože oba dva způsoby kódování označují stejné číslo 25, a to proto, že hodnotu 25 lze z číslice „25“ vyčíst přímo, aniž by bylo potřeba provádět nějaké dodatečné výpočetní operace, jako je tomu v případě notace „5²“. Explicitní reprezentování relevantních

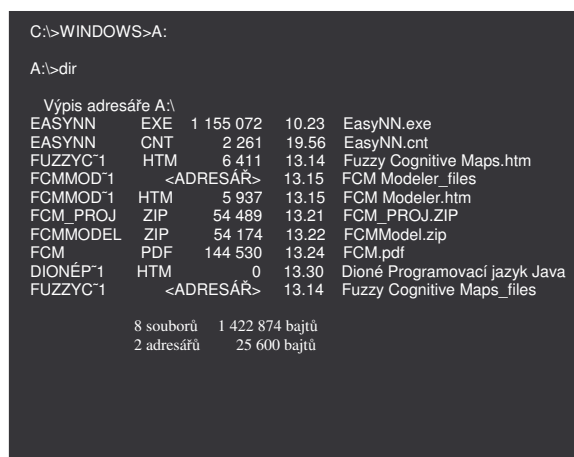
informací takto za člověka vykonává a realizuje velkou část usuzování a výpočtů, které jsou jinak nezbytným předpokladem úspěšného vyřešení problému. Ostatně na celý proces řešení problému se lze dívat jako na proces, kdy to, co původně bylo pouze implicitní, je postupně (prostřednictvím transformace a kombinování různých informací) činěno explicitním - „(vy)řešit problém jednoduše znamená reprezentovat tento problém takovým způsobem, aby jeho řešení bylo na první pohled zřejmé“¹ (Simon, 1981, s. 15).

Hlavní přínos kognitivních artefaktů tak tedy spočívá v tom, že transformují a zjednodušují problémovou doménu, ve které se musí lidská mysl pohybovat. Tento transformační efekt kognitivních artefaktů lidské mysli umožňuje, aby při řešení daného problému využívala svých silných stránek a zároveň přitom kompenzovala slabé stránky svého kognitivního profilu. Dennett (2004b) v této souvislosti hovoří o tzv. **kognitivních technologiích přereprezentování**, které převádějí komplexní informace o prostředí do jednodušších, přirozenějších a uživatelsky vstřícnějších formátů, ve kterých jsou tyto informace předložitelné vrozeným specializovaným perceptuálním a rozlišovacím schopnostem. Così ve stylu operačního systému Windows, který z ovládání počítače učinil záležitost manipulace s různými ikonami v prostředí počítačové obrazovky, což je činnost, která je z hlediska lidské mysli naprosto přirozená a intuitivní, neboť schopnost v prostředí rozpoznávat různé předměty a s těmito předměty různě manipulovat je člověku – s ohledem na evoluční historii, která stojí za současným „designem“ lidské mysli - do značné míry vrozená, počítačový odborník by řekl „natvrdo zadrátovaná“ (*hard-wired*). Tímto způsobem kognitivní artefakty převádějí nové problémy na starou mašinerii řešení, kterou má již lidská mysl dobře zvládnutou a to, co lidská mysl dobře zvládá, je především rozpoznávání různých obrazových a zvukových vzorců a struktur (*pattern-recognition*), sledování předmětů a manipulace s těmito předměty a s tím související senzomotorická koordinace; to, s čím má naopak lidská mysl obvykle problémy, jsou různé druhy sekvenčního myšlení jako je plánování nebo deduktivní či induktivní logické usuzování, kdy je potřeba provádět dlouhé posloupnosti operací s různými nenázornými entitami. Kognitivní vědec a filozof *Andy Clark* (2003, s. 5) tento kognitivní profil lidské mysli obrazně charakterizuje slovy „Dobrá ve frisbee, špatná na logiku“². V podobném duchu Cumminsová (1998, s. 196) konstatuje, že „nepočítáme-li řeč, jsme v oblasti symbolického uvažování bezúspěšně nemožní. Naproti tomu je člověk... na předních místech, pokud jde o rozpoznávání struktur a klasifikaci. [...] Cokoliv dalšího od nás vyžaduje velkou míru úsilí a značné spolehnutí se na abstraktní reprezentace.“ Tyto slabé a

¹ „Solving a problem simply means representing it so as to make the solution transparent.“

² „Good at Frisbee, Bad at Logic.“

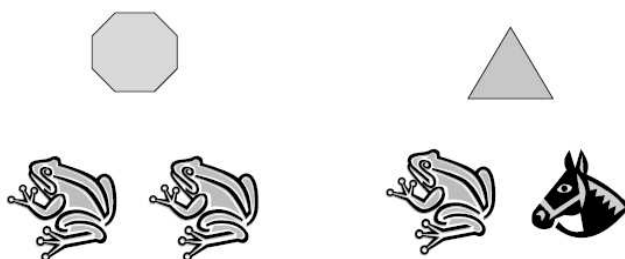
silné stránky lidské mysli byly svého času také ztělesněny ve dvou soupeřících počítačových prostředích *MacIntoshe* a *DOSu*, která italský sémiotik *Umberto Eco* (2000, s. 20) velice vtipně přirovnal ke katolictví, resp. k protestantismu. „*Jsem pevně přesvědčen, že MacIntosh je katolický a DOS protestantský. Vskutku je MacIntosh protireformační... Je radostný, přátelský, smířlivý, říká věřícím, jak musí postupovat krok za krokem, aby dosáhli, když už ne království nebeského, tak okamžiku, v němž se jejich dokument vytiskne. Je to katechetické: o podstatě zjevení se pojednává pomocí prostých formulací a nádherných vyobrazení. Každý má právo na spasení. DOS je protestantský nebo přímo kalvínský. Dovoluje svobodný výklad Písma, vyžaduje obtížná osobní rozhodnutí, nutí uživatele k subtilní hermeneutice a považuje za zaručenou myšlenku, že ne všichni mohou dosáhnout spasení. Aby systém fungoval, je třeba, abyste interpretovali program sami... Můžete namítnout, že s přechodem na Windows se vesmír DOS začal podobat blíže protireformační toleranci MacIntoshe.*“



Obrázek 14: Srovnání windowsovského prostředí ovlivněného „protireformační barokní okázalostí“ MacIntoshe (první obrázek) s „protestantskou strohostí“ DOSu (na druhém obrázku).

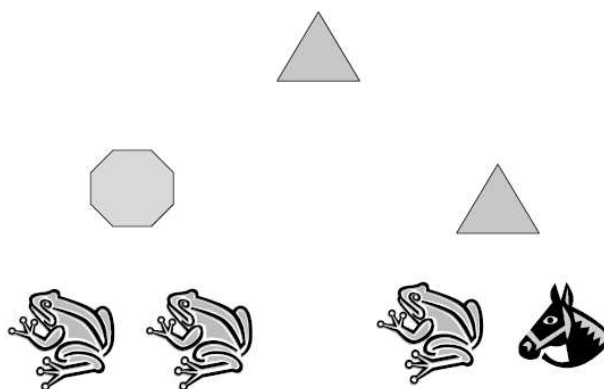
Základním transformačním efektem kognitivních artefaktů jako je abakus, logaritmické pravítko nebo i běžný aritmetický systém je to, že „platónské říši“ abstraktních objektů, vztahů a zákonitostí dávají konkrétní, viditelnou a hmatatelnou podobu, která je snadno přístupná lidskému senzomotorickému aparátu, prostřednictvím kterého člověk může jinak velice obtížně uchopitelnou doménu abstraktních čísel a číselných zákonitostí prozkoumávat a následně i využívat ve svých různých životních projektech. Rozdíl mezi výše uvedenými kognitivními artefakty je pouze ve způsobu implementace zákonitostí platících v abstraktním světě matematických objektů – tedy ve způsobu, jak jsou tyto objekty systematicky kódovány a jak je s daným typem symbolické reprezentace možné manipulovat tak, aby byla zachována

možnost smysluplné interpretace vstupně-výstupních vztahů odpovídající sémantice cílové domény čísel a aritmetických operací. Tyto kognitivní artefakty jsou součástí širší třídy **nástrojů mysli**, které vytvářejí prostřednictvím různých symbolických reprezentací zcela nové prostředí plné nových snadno identifikovatelných a manipulovatelných objektů, se kterými si dokáží dobře poradit i základní, nižší kognitivní procesy, které jsou spojené s percepcí, s rozpoznáváním a kategorizací objektů, s asociováním podnětů, s pamětí, učením či s představivostí. Tyto nástroje mysli fungují jako jakýsi zvláštní druh brýlí, které dokáží člověku zprostředkovat nové virtuální světy plné komplexních a abstraktních struktur a vztahů, které jsou pro člověka jinak neviditelné, nebo jenom velice obtížně přístupné. V této souvislosti byl proveden jeden velice zajímavý experiment, který názorně osvětluje roli nástrojů mysli při exploataci nových „kognitivních území“. Thompson a kol. (1997) ve svém výzkumu se šimpanzi (druhu *Pan troglodytes*) prokázali, že osvojení si jednoduchých arbitrárních symbolů označujících vztahy mezi předměty umožňuje šimpanzům rozpoznávat komplexní a abstraktní vztahy, které jsou jinak pro šimpanze bez této pomůcky nepřístupné. Šimpanzi se nejdříve učili spojovat vztah totožnosti (např. dvě žáby, dva koně) s jedním arbitrárním symbolem (plastovou značkou) a vztah odlišnosti (např. žába a kůň) s jiným arbitrárním symbolem (jinou plastovou značkou, viz obrázek 15). Poté se měli naučit

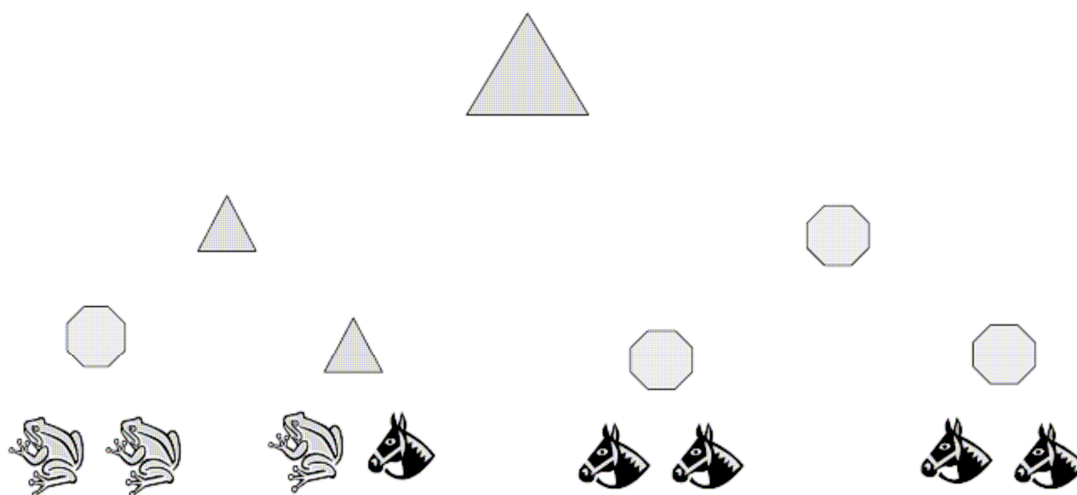


Obrázek 15: Nulá úroveň abstrakce.

identifikovat stejné vztahy i na vyšších, abstraktnější úrovních: Jestliže se na displeji objevily dva páry totožných předmětů (např. žába/žába a kůň/kůň), správnou reakcí šimpanze bylo posoudit vztah mezi těmito dvěma páry předmětů jako vztah totožnosti. Stejný vztah totožnosti platil i pro dva páry sestávající ze dvou odlišných předmětů (např. žába/kůň a pes/kočka). Kdyby se na displeji nacházel jeden pár totožných a jeden pár rozdílných předmětů, odpovídajícím vztahem (na první úrovni abstrakce) by byl vztah odlišnosti (viz obrázek 16). Podobným způsobem je možné pokračovat i na druhé a dalších úrovních abstrakce (viz obrázek 17).



Obrázek 16: *První úroveň abstrakce.*



Obrázek 17: *Druhá úroveň abstrakce.*

Šimpanzi, kteří nedostali šanci naučit se spojovat vztah totožnosti a odlišnosti se dvěma jednoduchými arbitrárními symboly (na rozdíl od těch, kteří tu šanci dostali), se nedokázali nikdy naučit identifikovat vztahy totožnosti a odlišnosti na abstraktnější úrovni, tedy na úrovni vztahů mezi vztahy (mezi vztahy...). Podle Thompsona a kol. je tento problém pro šimpanze, kteří si osvojili asociační spoje mezi jednotlivými typy vztahů a odpovídajícími plastovými značkami, zvládnutelný z toho důvodu, že při prezentaci každého páru předmětů se šimpanzovi v mysli vyvolá nějaká forma mentální reprezentace odpovídajícího plastického symbolu, díky čemuž se celý problém do značné míry zredukuje z abstraktního na názorný („Jsou ty dva mentální obrazy, které se mi vybavily, totožné, nebo nejsou?“).

Obdobným způsobem je u člověka veškerá zkušenost s většími čísly a se složitějším počítáním primárně zprostředkována fyzikální strukturou nějakého smyslově vnímatelného reprezentačního systému (Lakoff, Núñez, 2000). Díky častému zacházení s daným

reprezentačním systémem si pak člověk časem dokáže fyzikální strukturu takového systému celkem dobře představit i v situacích, kdy není aktuálně ve vnějším prostředí přítomna; stejným způsobem je možné se naučit ve své představivosti systematicky manipulovat s představovanou fyzikální strukturou reprezentačního systému a provádět tímto způsobem požadovaný výpočet tzv. „z hlavy“. Tímto se to, co mělo původně podobu vnější, fyzikálně implementované symbolické reprezentace, se kterou člověk manipuluje prostřednictvím svých senzomotorických dovedností, stává pouze představovanými reprezentačními strukturami, se kterými člověk manipuluje pouze ve své představivosti (Rumelhart a kol., 1986a).

Přesto se člověk spoléhá na vnější strukturu reprezentačního systému de facto při každém jenom trochu složitějším výpočtu, který je náročnější na kapacitu pracovní paměti. Klasickým příkladem je zde již jednou zmiňované násobení velkých čísel, při němž si člověk většinou pomáhá tím, že počítá s pomocí tužky a papíru: *„Poznámkami komunikuje... mozek sám se sebou. Je pro něj mnohem snazší zpracovávat... problémy a informace na základě poznámek na papíře, než je udržovat „ve vzduchu“, v sobě samém.“* (Buzan, 2002, s. 33) *„Počítání na papíře má [však] více užitečných funkcí; [kromě toho, že] poskytuje spolehlivé úložiště pro mezivýsledky,... jednotlivé symboly slouží i jako značky, které lze následovat, a jak naše oči a prsty dospívají na kterékoli místo, připomínají nám, jaký krok má v tomto více než dobře známém předpisu následovat.“* (Dennett, 2004b, s. 129)

Tento způsob počítání je typem obecnějšího, tzv. **stigmergického algoritmu** (Clark, 1997), který představuje specifický způsob řízení a koordinace komplexní posloupnosti činností, který není založen na nějakém centrálním plánu, který by detailně specifikoval každý jednotlivý krok, ale na „lokálních“ pravidlech, která specifikují relativně jednoduchou reakci na přítomnost určité struktury ve vnějším prostředí; důsledkem takové reakce je modifikace stávajících nebo vznik nových struktur ve vnějším prostředí; takto změněný stav vnějšího prostředí pak slouží jako spouštěč další reakce, která má za následek další modifikaci vnějšího prostředí, které ve své pozměněné podobě slouží jako vstup do dalšího kola stigmergického algoritmu. Odtud je také odvozen název celé procedury – *stigmergický* jako kombinace slov pro *znak (stigma)* a *práci (ergon)*, tj. využívání produktů práce jako signálů pro vykonávání další práce. Dobře známým příkladem využití tohoto algoritmu v každodenním životě je jednoduchá mnemotechnická pomůcka: Řekněme, že se budu snažit nezapomenout vynést odpadky. Abych trochu ulehčil své paměti, dám jednoduše pytel s odpadky přede dveře, takže až budu vycházet ven, „praští“ mě pytel s odpadky do očí a okamžitě mi připomene, co mám udělat. Tímto zásahem do vnějšího prostředí – postavením

pytle s odpadky přede dveře - jsem v něm vytvořil určitou strukturu, která funguje jako signál, kterým komunikuji sám se sebou a s jehož pomocí organizuji a řídím své chování tak, abych dosáhl cílů, které jsem si předsevzal. Podle Dennetta (2004b) člověk tímto způsobem běžně přesouvá část úkolu sebe-řízení na vnější prostředí, ve kterém vytváří periferní systémy uchovávání informací v podobě různých značek, signalizátorů a jiných náznaků. Člověk takto nechává na vnějším světě, aby uchovával nějakou snadno převoditelnou informaci, takže kapacitu své mysli si potom může šetřit na něco důležitějšího. Toto přenášení kognitivní zátěže na předměty vnějšího světa, které hrají aktivní roli v řízení různých kognitivních procesů, je podle některých kognitivních vědců nedílnou součástí toho, co dělá člověka člověkem. Takto například Dennett (2004b, s. 125-126) tvrdí, že *„zdrojem naší vyšší inteligence... je... přesouvání maxima našich kognitivních úkolů na prostředí samo – vylévání našich myslí (tj. našich mentálních projektů a aktivit) do okolního světa, kde mohou být naše významy skladovány, zpracovávány a přereprezentovávány spoustou periferních zařízení, které konstruujeme, a tím jsou vyhlazovány, vylepšovány a ochraňovány ty procesy transformování, které jsou naším myšlením. Tato rozšířená praxe přesouvání [...] zmenšuje kognitivní zatížení našeho vnímání a naší paměti [a] osvobozuje nás z mezí našich živočišných mozků.“* Podle Andyho Clarka a Davida Chalmerse (1998) vzniká interakcí lidské mysli s externím médiem **hybridní kognitivní systém**, jehož všechny komponenty, vnitřní i vnější, hrají aktivní kauzální roli v řízení a generování chování, takže odstraníme-li vnější komponentu, behaviorální kompetence celého systému se zhroutí úplně stejně, jako kdybychom odstranili nějakou část mozku. Clark (2001) ilustruje tuto skutečnost na příkladu starých lidí, kteří jsou často zvláštním způsobem zneschopněni, když jsou přesunuti ze svých domovů do nemocnic nebo do nějakého jiného, pro ně cizího prostředí: *„starí lidé, kteří jsou ze svých domovů přesunuti do nemocnic,... získávají obrovský hendikep... Často se zdají být naprosto dementní – naprosto neschopní jíst, oblékat se a mýt se, o zajímavějších činnostech ani nemluvě. Jsou-li ale navráceni do svých domovů, jsou často schopni se o sebe docela dobře postarat. Jak to dělají? Po léta své domácí prostředí naplňovali dokonale známými značkami, spínači pro zvyky, připomínkami, co dělat, kde najít jídlo, jak se obléci, kde je telefon atd. Stará osoba může být v takovém více než dobře známém světě doslova virtuosem svépomoci, navzdory tomu, že se jeho nebo její mozek stává stále neschopnější nových kol učení... Vztít je z jejich domovů znamená doslova oddělit je od velké části jejich mysli – je to něco, co je potenciálně tak zničující jako podstoupit operaci mozku.“* (Dennett, 2004b, s. 127-128) Clark (2003), aby zdůraznil, jak hlubokou a integrální součástí lidských systémů řešení problémů jsou různé kognitivní technologie, mluví o jednotlivých kognitivních technologiích

jako o **mindwarových upgradech** a o člověku pak jako o **kognitivním hybridu** nebo dokonce jako o **(při)rozeném kyborgovi**, jehož mysl je distribuována jak v rámci biologického mozku tak i v rámci různých nebiologických „obvodů“. Tuto pro většinu lidí jen těžko přijatelnou myšlenku Clark spolu s Chalmersem (1998) ilustrují pomocí analogie s rybou, která je podle nich tak výtečným plavcem částečně proto, že dokáže propojit své plovací pohyby s vnějšími zdroji kinetické energie, které se v jejím vodním prostředí nachází v podobě různých vírů a turbulencí. Tyto víry a turbulence zahrnují jak ty přirozeně se vyskytující (např. když do vody spadne nějaký kámen) tak i ty, které si ryba sama vytváří (pomocí dobře načasovaného mávnutí ocasem či ploutví). Ryba pak plave tak, že vestavuje tyto ve vnějším světě se vyskytující procesy do svých vlastních lokomočních procedur, takže ryba a okolní turbulence vytváří propojený hybridní a vysoce efektivní plovací systém.

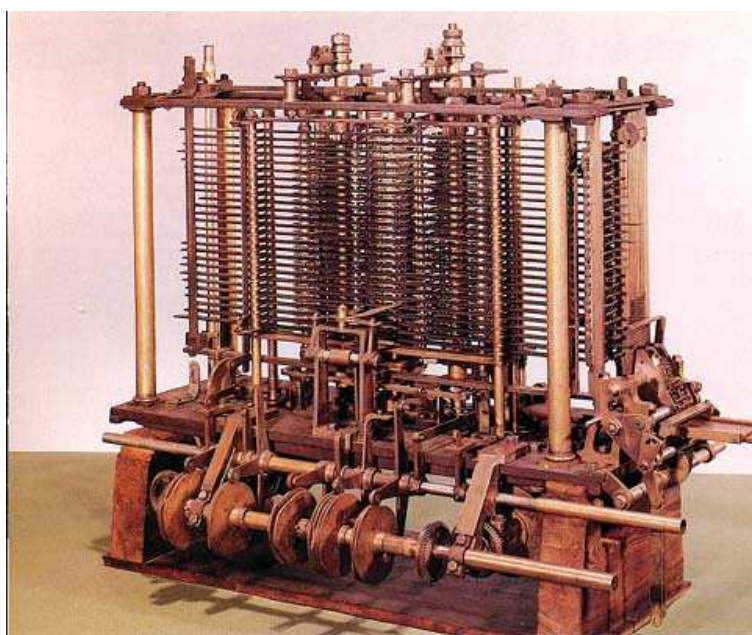
Výše uvedený příklad s mnemotechnickou pomůckou pro zapamatování a realizaci plánovaných činností je jen jednoduchou ilustrací možností stigmergického algoritmu. S jeho pomocí je však možné řídit a koordinovat rovněž vysoce komplexní posloupnosti činností jakou je například ta, kterou zahrnuje navigace vojenské námořní lodě o výtlačku několika set tisíc tun (Hutchins, 1995): Navzdory tomu že samotný proces navigace představuje jako celek vysoce komplexní výpočet, díky specifické dělbě (kognitivní) práce není zapotřebí žádného centrálního koordinátora nebo plánovače, který by měl přehled o celém výpočetním procesu a který by na základě této své znalosti určoval jednotlivým členům navigačního týmu, co mají v určitou chvíli dělat. Ve skutečnosti členové navigačního týmu nedělají nic jiného než že čekají na určité specifické změny/signály ve svém prostředí (například v podobě slovního rozkazu, zvonění zvonku, dodání tabulky s číselnými údaji potřebnými pro nějaký relativně jednoduchý výpočet atd.), aby mohli zahájit svoji poměrně jednoduchou a omezenou činnost, která vede k modifikaci prostředí jiných členů navigačního týmu, kteří zase na základě této změny spouští své vlastní jednoduché vzorce chování ovlivňující chování dalších členů navigačního týmu atd. dokud není celý problém (například určení pozice lodi před vstupem do přístavu) vyřešen.

Uvědomění si této možnosti rozložit komplexní numerický výpočet na soubor jednoduchých hierarchicky uspořádaných matematických operací stálo podle Gigerenzera (2000) u zrodu prvního programovatelného mechanického počítače, který v polovině 19. století sestavil anglický matematik a vynálezce *Charles Babbage*. Babbageovým cílem bylo nahradit strojem „počítače“, což v Babbageově době byli lidé, kteří prováděli výpočty nutné k sestavení logaritmických tabulek usnadňujících matematikům, astronomům či navigátorům jejich vlastní výpočty. To, co Babbageovi umožnilo přijít s myšlenkou nahradit lidský počítač

počítačem mechanickým, bylo oddělení numerického výpočtu od zbytku lidské inteligence. V osvíceneckém 18. století se schopnost počítat považovala za výsostný znak lidské racionality a samotné lidské myšlení bylo nahlíženo jako jistý druh kombinatorického výpočtu, ve smyslu kombinace jednotlivých dojmů (počiteků a vjemů) a idejí (obrazů dojmů) (Störig, 1993), což je myšlenka, na kterou v polovině 20. století svým způsobem navází *Allen Newell* a *Herbert Simon* v rámci své koncepce fyzikálního symbolového systému (*physical symbol system hypotheses*) jako nutné a dostačující podmínky pro vytvoření systému vykazujícího obecnou inteligenci (více viz oddíl „1.2.3.2 Symbolové systémy“). Od přelomu 18. a 19. století se však začalo na numerické počítání pohlížet jako na veskrze monotónní, rutinní a neustále dokola se opakující činnost, která od člověka nevyžaduje příliš představivosti ani příliš inteligence, takže ji může vykonávat i nepříliš kvalifikovaná pracovní síla. K této změně pojetí povahy numerického výpočtu přispělo významným způsobem zavedení principu dělby (kognitivní) práce do procesu „výroby“ různých astronomických a navigačních tabulek. Takto například ke konci 18. století dostal inženýr *Gaspard Riche de Prony* od francouzské vlády za úkol zorganizovat projekt, jehož cílem bylo vypočítat 10 000 sinusových hodnot s přesností na 25 desetinných míst a 200 000 logaritmů s přesností na 14 nebo 15 desetinných míst. Při řešení této zakázky se Prony nechal inspirovat způsobem výroby v tehdejších manufakturách, ve kterých byl celý výrobní proces rozložen na řadu dílčích, relativně jednoduchých pracovních úkonů. Prony podobným způsobem rozložil celý projekt na řadu hierarchicky uspořádaných úkolů a podúkolů: Na vrcholu celé hierarchické pyramidy bylo několik málo profesionálních matematiků, kteří byli zodpovědní za navrhování výpočetních procedur umožňujících výpočet požadovaných výsledků. V současné terminologii bychom je mohli nazvat projektovými manažery, kteří předem určují, co, kdo, kdy a jak bude dělat, aby se dosáhlo stanovených cílů. Uprostřed pak bylo 7 až 8 vyškolených analytiků, kteří zpracovávali údaje, které jim ze samého dna hierarchické pyramidy dodávalo 70 až 80 nekvalifikovaných pracovníků, kteří ovládali pouze základy aritmetiky a kteří neustále dokola prováděli tisíce a tisíce jednoduchých výpočetních operací jako je sčítání nebo odčítání (Gigerenzer, 2000). Podobný lidský počítač vznikl rovněž za druhé světové války v rámci projektu Manhattan v Los Alamos, kde bylo při práci na konstrukci atomové bomby potřeba provádět řadu komplikovaných výpočtů. Vzhledem k omezeným možnostem tehdejší výpočetní techniky, která dokázala pouze sčítat, odčítat, násobit a s určitými problémy i dělit, byli tamější pracovníci nuceni sestavit lidský počítač, který se skládal z osob, které opakovaně prováděli jednoduché výpočty a jejich výsledky odesílali dalším

osobám, které ho použily jako vstup do své vlastní jednoduché výpočetní procedury atd. (Gigerenzer, 2000).

Když se ukázalo, že komplexní výpočet může být realizován armádou nekvalifikovaných dělníků, mohl Babbage začít uvažovat o tom, jak tyto dělníky - vykonávající neustále dokola se opakující jednoduché výpočty - nahradit strojem. Výsledkem jeho snahy byl *Analytický stroj* (*Analytical Engine*, viz obrázek 18), jehož architektura se v mnohém podobala struktuře dobových manufaktur a továren. Když Babbage popisoval jednotlivé části svého Analytického stroje, hovořil o tzv. „mlýnici“ (*mill*), kde se realizovaly aritmetické výpočty (obdobu procesoru dnešních počítačů), a o tzv. „skladišti“ (*store*), kde se skladovala již zpracovaná čísla nebo čísla, která měla být teprve zpracována (ekvivalent počítačové paměti). Tento metaforický popis měl svojí zdrojovou oblast ve způsobu, jakým se v té době pracovalo v textilních továrnách, ve kterých se příze nejdříve vzala ze skladiště do mlýnice,



Obrázek 18: Babbageův Analytický stroj (*Analytical Engine*).

kde byla utkána v látku, která byla následně odeslána zpátky do skladiště. V Analytickém stroji se stejným způsobem nejdříve dopravila čísla ze „skladiště“ do aritmetické „mlýnice“ ke zpracování, jehož výsledek se následně odeslal zpátky do „skladiště“. Analytický stroj takto vyráběl čísla obdobným způsobem, jakým tradiční manufaktury a továrny té doby vyráběly běžné zboží.

1.2.1.3 Formální logika

Všechny zde zatím uvedené příklady formálních systémů byly z oblasti matematiky - všechny tedy tím či oním způsobem modelovaly některé z prvků a zákonitostí vyskytujících se v doméně matematického myšlení. Přestože je matematické myšlení důležitou součástí lidské kognice, v každodenním životě člověka se jeho racionalita projevuje především tzv. „zdravým selským rozumem“ (*common sense*), tedy schopností analyzovat a interpretovat situace na základě kontextu za použití rozsáhlé databáze navzájem propojených a k sobě odkazujících poznatků o světě a o tom, jak to v něm chodí. Tato rozsáhlá databáze poznatků pak ve spojení s určitými inferenčními pravidly člověku umožňuje dostat se na základě určitých bezprostředně dostupných údajů k informacím, které bezprostředně dostupné nejsou (Bruner, 1957). Člověk například může být na základě hodin přírodopisu na základní škole majitelem přesvědčení, že *granát patří mezi polodrahokamy*; takovému člověku se pak může stát, že při rytí své zahrady narazí na kámen, o kterém si bude myslet, že by to mohl být granát, ale protože si tím nebude jistý, půjde se zeptat odborníka a ten mu řekne, že *tento kámen určitě není polodrahokam*; na základě zbytků svých chabých znalostí z mineralogie a na základě informace od odborníka si člověk může okamžitě udělat správný závěr, že *kámen, který našel na své zahradě, určitě není granát*. Jindy se zase člověk může na nějaké společenské akci například dozvědět, že jeho přímý nadřízený, říkejme mu *Petr*, je *ženatý buď s Marií, anebo s Pavlou*; doma se potom od své lépe informované manželky dozví, že *Petr určitě není ženatý s Pavlou*, protože ta je vdaná za Honzu; má-li člověk k dispozici všechny tyto informace, může potom učinit závěr, že *Petr tedy musí být ženatý s Marií*.

Tyto „zákony lidského myšlení“ - tj. inferenční pravidla, kterými se člověk řídí při přechodech mezi svými jednotlivými přesvědčeními - se snaží v rámci své abstraktní symbolické notace zachytit, popsat a modelovat formální logika. Například usuzování o vztahu mezi Petrem, Pavlou a Marií z výše uvedeného příkladu lze v rámci notace predikátové logiky zachytit prostřednictvím následující sady výrazů, které vedle abstraktního schématu zachycujícího vztah mezi dvěma předměty:

[vztah mezi prvky] ([prvek představující *subjekt*], [prvek představující *objekt*])

obsahují rovněž dva logické operátory *nebo* (\vee) a *negace* (\neg):

- (1) [Ženatý] ([Petr], [Marie]) *nebo* [Ženatý] ([Petr], [Pavla])
- (2) *negace* [Ženatý] ([Petr], [Pavla])
- (3) tedy [Ženatý] ([Petr], [Marie])

Důležité je, že (3) vyplývá z (1) a (2) bez ohledu na to, jaký je konkrétní sémantický obsah jednotlivých výrazů, takže dané inferenční pravidlo bychom mohli zapsat i takovým způsobem, který se vůbec nezmiňuje o manželství, Petrovi, Pavle nebo Marii nebo o čemkoli, co má co do činění se světem reálných předmětů a událostí. V rámci symboliky výrokové logiky můžeme tyto výrazy nahradit například písmeny p a q a inference přitom bude i nadále fungovat:

- (1) $p \vee q$
- (2) $\neg q$
- (3) tedy p

Podobně lze zapsat i usuzování našeho mineraloga-amatéra (mající logickou strukturu nazývanou *modus tollens*):

- (1) $p \rightarrow q$
- (2) $\neg q$
- (3) tedy $\neg p$

Obdobným způsobem je teoreticky možné zachytit logickou strukturu všech inferenčních procesů, které tvoří podstatu lidské racionality, a to bez ohledu na sémantický obsah „materiálu“, na který jsou inferenční procesy aplikovány, neboť výroková logika abstrahuje v lidském myšlení od všeho, co není důležité z hlediska vyplývání, tj. z hlediska těch inferenčních zákonitostí, které zabraňují tomu, aby z pravdivosti vyplývala nepravdivost. To, co je z hlediska vyplývání relevantní, je pouze pravdivostní hodnota jednotlivých výroků (*pravda/nepravda*, $1/0$) a logická syntaxe, která prostřednictvím logických operátorů (*a*, *nebo*, *negace*) spojuje jednoduché výroky ve výroky složené. Ve výrokové logice přitom nejsou důležité konkrétní výroky, ale vždy jen *výroková forma*, která se skládá z proměnné za výrok a z logického operátoru. Poté, co jsou do výrokové formy dosazeny výroky s konkrétní pravdivostní hodnotou, lze na základě znalosti použitého logického operátoru (a znalosti pravdivostních hodnot jednotlivých částí složeného výroku) určit pravdivostní hodnotu celého složeného výroku. Logický operátor je tak *pravdivostní funkcí*, která pravdivostním hodnotám

částí přiřazuje jednoznačnou pravdivostní hodnotu celku (Sousedík, 1999). Tyto pravdivostní funkce je možné reprezentovat v podobě tzv. *pravdivostních tabulek logických operátorů*.

p	q	$p \wedge q$	p	q	$p \vee q$	p	$\neg p$
1	1	1	1	1	1	1	0
1	0	0	1	0	1	0	1
0	1	0	0	1	1		
0	0	0	0	0	0		

a
b
c

Tabulka 3a, b, c: Pravdivostní tabulky pro dvojmístné pravdivostní funkce „a“, „nebo“ a „negace“. a) „ $p \wedge q$ “ je pravdivé pouze a jenom tehdy, když „ p “ je pravdivé a „ q “ je pravdivé. Ve všech ostatních případech je „ $p \wedge q$ “ nepravdivé. b) „ $p \vee q$ “ je pravdivé, jestliže je pravdivé „ p “, nebo „ q “, nebo „ p “ i „ q “. „ $p \vee q$ “ je nepravdivé pouze tehdy, kdy jsou „ p “ i „ q “ nepravdivé. c) Logický operátor negace invertuje pravdivostní hodnotu „ p “ v její opak.

Výše uvedené tabulky pravdivostních funkcí jsou de facto totožné s tabulkami, které vyjadřují nějakou matematickou funkci, např. x^2 . Jediný rozdíl mezi nimi spočívá v tom, že zatímco matematické funkce operují s čísly (tzn. že číselným vstupům přiřazují určité číselné výstupy), pravdivostní funkce kalkulují s pravdivostními hodnotami (tzn. že pravdivostním hodnotám částí přiřazuje pravdivostní hodnotu celku).

V souladu s touto těsnou paralelou mezi matematickými a pravdivostními funkcemi se anglický logik z 19. století *George Boole* ve své knize *An Investigation of the Laws of Thought* (Zkoumání zákonů myšlení) pokusil zachytit logiku lidského myšlení prostřednictvím jednoduchých „matematických“ operací definovaných na pravdivostních hodnotách výroků reprezentovaných číslicemi 1/0: Boole vytvořil speciální jazyk – tzv. *booleovskou algebru* - k popisu výroků a různých způsobů zacházení s těmito výroky, který umožňuje určovat pravdivostní hodnotu i vysoce komplexních výroků. Booleovská algebra - podobně jako číselná algebra – sestává z několika konstant, proměnných a základních operací. Zatímco v číselné algebře je mnoho možných konstant (v podobě všech celých čísel a zlomků), v booleovské algebře existují pouze dvě možné konstanty (pravda/nepravda), arbitrárně reprezentované číslicemi 1/0, které v tomto případě neoznačují žádné číslo nebo množství, jak tomu u číslic obvykle bývá, ale mají čistě rozlišující funkci - stejně dobře by posloužila i označení Petr/Pavel nebo ♠/♣; jedná se de facto o označení základní jednotky informace, **bitu**, který – slovy filozofa Gregoryho Batesona – označuje **nejmenší možný rozdíl, který znamená rozdíl**, tj. rozdíl, který dělí sledované proměnné do dvou jednoznačných tříd (kterými jsou v případě booleovské algebry pravda/nepravda). Proměnné v booleovské

algebře označují výroky, které mohou být buď pravdivé, nebo nepravdivé; booleovské proměnné tedy mohou nabývat dvou různých (pravdivostních) hodnot (1/0). Tyto proměnné se v booleovské algebře označují velkými tiskacími písmeny (A, B, C...). Posledním prvkem v rovnicích booleovské algebry jsou operace, které umožňují kombinovat a transformovat existující veličiny na veličiny nové. V číselné algebře jsou základními operacemi sčítání, odčítání, násobení a dělení; v booleovské algebře to jsou operace inverze, logického násobení a logického sčítání, které ve výrokové logice odpovídají logickým operátorům „negace“ (\neg), „a“ (\wedge), resp. „nebo“ (\vee) (Hoernes, Heilweil, 1969). V tabulce 4 jsou uvedeny definice těchto základních logických operací a také několik nejdůležitějších pravidel pro manipulaci a transformaci veličin booleovské algebry.

Logické násobení	Logické sčítání	Logická inverze
$0 \cdot 0 = 0$	$0 + 0 = 0$	$\overline{1} = 0$
$0 \cdot 1 = 0$	$0 + 1 = 1$	$\overline{0} = 1$
$1 \cdot 0 = 0$	$1 + 0 = 1$	
$1 \cdot 1 = 1$	$1 + 1 = 1$	
<ul style="list-style-type: none"> ■ Násobit a sčítat lze tři i více konstant: Nejprve se spolu vynásobí/sečtou jakékoli dvě konstanty, výsledek se potom vynásobí/sečte s třetí konstantou atd. ■ Pořadí operací lze vyznačit s pomocí závorek; přitom platí, že přednost mají operace zapsané uvnitř závorek před operacemi stojícími mimo závorky ■ Operace inverze se vždy provádí před operacemi násobení a sčítání ■ Operace se vždy provádí v pořadí 1) inverze, 2) násobení, 3) sčítání ■ Logické násobení i sčítání je komutativní: $A \cdot B = B \cdot A$ $A + B + C = B + A + C$ ■ Logické násobení i sčítání je asociativní: $A \cdot (B \cdot C) = (A \cdot B) \cdot C$ $A + (B + C) = (A + B) + C$ 		

Tabulka 4: Definice booleovských operací logické inverze, logického násobení a logického sčítání.

S pomocí tohoto symbolického aparátu je relativně snadné určit pravdivostní hodnotu i velice složitých výroků nebo dokázat rovnost dvou výroků v booleovské rovnici, například $A + A \cdot B = A$ (viz tabulka 5).

A	B	$A \cdot B$	$A + A \cdot B$
0	0		
0	1		
1	0		
1	1		

\Rightarrow

A	B	$A \cdot B$	$A + A \cdot B$
0	0	0	
0	1	0	
1	0	0	
1	1	1	

\Rightarrow

A	B	$A \cdot B$	$A + A \cdot B$
0	0	0	0
0	1	0	0
1	0	0	1
1	1	1	1

Tabulka 5: Pravdivostní tabulky dokazující rovnost dvou výroků ve větě $A + A \cdot B = A$.

Avšak to nejdůležitější, co vyplývá z Booleovy práce, je „*demonstrace překvapivé síly a obecnosti několika [málo] jednoduchých logických operací*“. (Hillis, 2002, s. 16). **Logické operace „a“, „nebo“ a „inverze“** totiž představují jakousi **univerzální sadu nástrojů, které umožňují sestavit jakékoli pravidlo pro transformaci vstupů na výstupy**. Takovým trochu složitějším, ale stále ještě dost jednoduchým pravidlem pro tvorbu výstupů ze vstupů je například „hlasovací“ funkce, která umožňuje třem vstupům „hlasovat“ o výstupu. Pravdivostní tabulka takové funkce je zachycena v tabulce 6.

Vstupy			Výstup
A	B	C	X
0	0	0	0
0	0	1	0
0	1	0	0
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	1
1	1	1	1

Tabulka 6: Pravdivostní tabulka „hlasovací“ funkce.

Jeden z možných způsobů implementace této hlasovací funkce prostřednictvím booleovských operací logického sčítání a násobení zachycuje výrok $(A \cdot B) + (A \cdot C) + (B \cdot C)$. Úspěšnost této implementace je možné ověřit dokázáním rovnosti dvou výroků v booleovské rovnici $(A \cdot B) + (A \cdot C) + (B \cdot C) = X$ (viz tabulka 7).

Vstupy							Výstup
A	B	C	A . B	A . C	B . C	$(A \cdot B) + (A \cdot C) + (B \cdot C)$	X
0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0
0	1	0	0	0	0	0	0
0	1	1	0	0	1	1	1
1	0	0	0	0	0	0	0
1	0	1	0	1	0	1	1
1	1	0	1	0	0	1	1
1	1	1	1	1	1	1	1

Tabulka 7: Pravdivostní tabulka dokazující rovnost dvou výroků v rovnici $(A \cdot B) + (A \cdot C) + (B \cdot C) = X$.

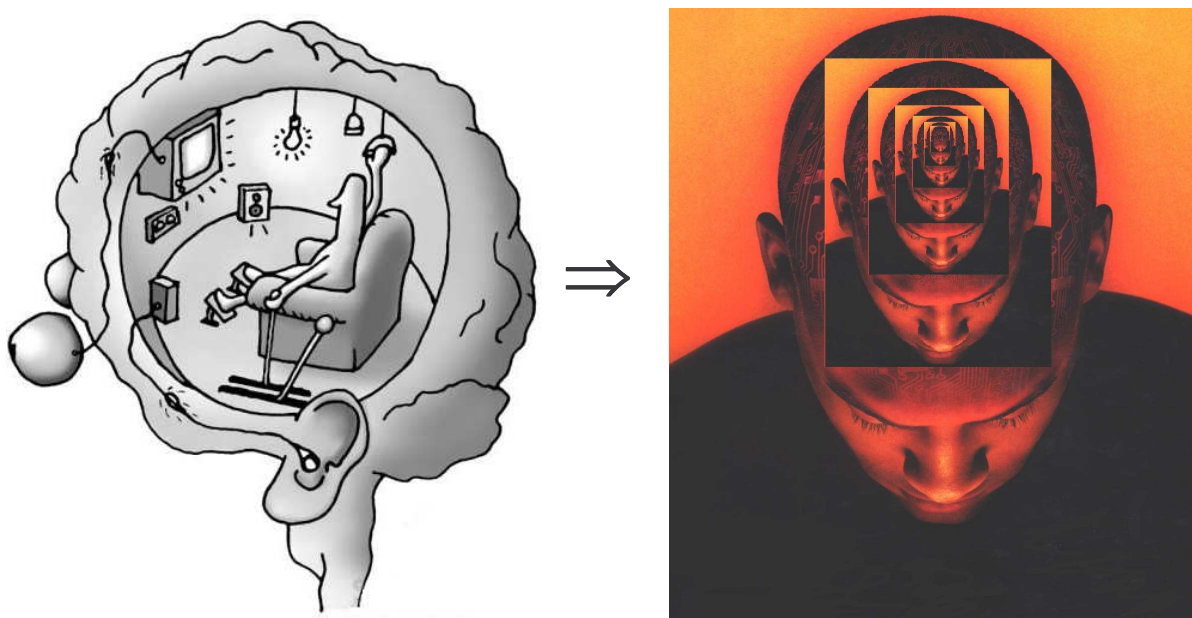
Z výše uvedeného popisu některých prvků formální logiky je zřejmé, že formální logika představuje systém, který se skládá a) ze souboru symbolů, b) z pravidel, která určují, jakým způsobem je možné jednotlivé symboly spojovat k vyjádření různých komplexních výroků, a

nakonec c) z pravidel specifikujících povolené způsoby vyvozování nových symbolů z těch starých. Klíčovou vlastností formální logiky je přitom to, že když se člověk pevně drží těchto pravidel, má jistotu, že se mu nikdy nestane, aby vyvodil nepravdivý závěr z pravdivých premis, a to přesto, že nebude mít žádnou konkrétní představu o sémantice symbolů, tedy o tom, co jednotlivé symboly reprezentují - stačí se držet pravidel, a „zákon o zachování pravdy“ se už o vše postará. Formální logika je tak formou hry – podobně jako šachy -, která se hraje s významu-prostými symboly podle pravidel, která se vztahují pouze k fyzikálním charakteristikám jednotlivých symbolů, nikoli k jejich významům. Podle Clarka (2001) to celé to trochu připomíná situaci člověka, který není příliš zběhlý v manuální činnosti, který je ale přesto schopen sestavit si nábytek na základě písemných instrukcí v návodu, který ho krok za krokem navádí, jak má jednotlivé dílce různých tvarů a velikostí smontovat dohromady. Vnějšímu pozorovateli se může zdát, že jednání takového člověka je zakotveno v hlubokém porozumění principům a zákonitostem práce se dřevem, ve skutečnosti se ale takový člověk pouze slepě drží instrukcí, které mu říkají, jaké pohyby má kdy a s čím vykonat. Formální logika je tak dalším příkladem formálního systému, který umožňuje implementovat sémantické charakteristiky lidského myšlení ve světě, jehož fungování je na sémantice lidského myšlení zcela nezávislé; tento svět je tvořen různými znaky a symboly, které je možné rozeznat pouze na základě jejich fyzikálních charakteristik a které jsou manipulovány a transformovány čistě na základě formálních/syntaktických pravidel, která jsou citlivá pouze k fyzikální podobě těchto znaků a symbolů. Přesto chování tohoto světa dokáže odrážet zákonitosti platící ve světě významů a racionálního usuzování. Výstižně to formuloval John Haugeland (1981, s. 23), když napsal „*If you take care of the syntax, the semantics will take care of itself*“, tedy jestliže se postaráme o syntax, sémantika se už o sebe postará sama.

Tento poznatek - že sémantiku lidského mentálního života je možné implementovat prostřednictvím jednoduchých manipulací s reálnými fyzickými objekty, a to na základě pravidel, která odkazují pouze k formě objektů, se kterými je manipulováno, a že entity, o kterých se předpokládá, že existují uvnitř lidské mysli (mentální reprezentace), je možné modelovat pomocí entit, které se nacházejí mimo lidskou mysl (tj. pomocí symbolických reprezentací nějakého formálního systému) – byl prvním krokem k výpočetní teorii mysli, která si činí nárok odpovědět na otázku, jakým způsobem je možné fyzicky realizovat lidskou racionalitu. Druhým krokem k výpočetní teorii mysli byla konstrukce automatických formálních systémů.

1.2.2 Automatické formální systémy

Všechny zde zatím uváděné příklady formálních systémů vyžadovali nějakého lidského operátora, schopného - na základě (syntaktických) pravidel transformace různých symbolických reprezentací - fyzicky realizovat (zákonitosti lidského mentálního života zachycující) přechody mezi jednotlivými (syntaktickými) stavy formálního systému (které ovšem mají po náležitě interpretaci i svoji sémantiku). Jestliže ale chceme lidskou racionalitu prostřednictvím takového procesu vysvětlit, potom skutečnost, že k jeho realizaci potřebujeme lidského operátora, představuje dosti závažný problém, neboť ve svém vysvětlení předpokládáme to, co chceme vysvětlit, tzn. racionálního lidského operátora. Ve filozofii mysli je tento problém znám pod názvem **karteziánské divadlo**, kterým Daniel Dennett (1993) označuje přirozenou tendenci lidí - včetně mnoha vědců a filozofů postulujících různé teorie mysli nebo dílčích kognitivních procesů¹ - předpokládat existenci nějakého privilegovaného místa v lidském mozku, jakéhosi **homunkula**, malého inteligentního človíčka, ke kterému se z periferie ke zpracování sbíhají všechny informace, na základě kterých pak řídí své tělo podobně jako loutkář vede a řídí svou loutku. Problém teorií (často nevědomě) vycházejících z takového předpokladu je v tom, že nic nevysvětlují, ale vysvětlení pouze odsouvají a odkládají - ve svém důsledku totiž vedou k nekonečnému



Obrázek 19a: Karteziánské divadlo a z něj vyplývající riziko nekonečného regresu malých homunkulů.

¹ Příznačná je v tomto ohledu Fodorova teorie modulární mysli (Fodor, 1983), která předpokládá existenci několika (automatických, mechanických, rychlých, nevědomých, „povinných“ a „informačně zapouzdřených“) periferních vstupních modulů, které zpracovávají vstupní senzorické informace a následně je odesílají k dalšímu zpracování centrálním kognitivním procesům (které jsou pod vědomou kontrolou, „informačně nezapouzdřené“, „nepovinné“, pomalé... a které tvoří podstatu specificky lidského myšlení).

regresu malých a stále menších homunkulů, kteří si v nekonečné řadě od největšího k nejmenšímu jako horký brambor předávají zodpovědnost za vysvětlení daného mentálního jevu. K takové nekonečné sérii homunkulů by vedla například teorie zrakového vnímání, která by předpokládala, že výsledkem zrakového vnímání je nějaký obraz; takový obraz by však vždy potřeboval někoho, kdo by se na něj díval, čímž se dostáváme zpátky k tomu, co jsme chtěli naší teorií původně vysvětlit. Stejně tak i při explanaci myšlení jako transformace a manipulace různých symbolických reprezentací hrozí nekonečný regres podobných homunkulů, kteří by se tentokrát nedívali na obrazy, ale na různé symbolické reprezentace, se kterými by podle určitých pravidel různě manipulovali.

Řešení problému s karteziánským divadlem a s nekonečnou regresí homunkulů spočívá ve využití principu **dělby práce a stigmergického algoritmu** (viz s. 37), kdy se veškerá práce rozděluje mezi ohromné, ale konečné množství hierarchicky uspořádaných homunkulů (nazývaných také **agenty** či **démoni**), kteří nikdy neduplikují schopnosti a dovednosti celého systému (jehož chování se má vysvětlit), ale vykonávají pouze omezené akce, spočívající v jednoduchém reagování na změny v prostředí tvořeném dalšími homunkuly reagujícími na reakce jiných homunkulů, kteří zase reagují na reakce dalších homunkulů atd. Tradičním organizačním schématem souboru takových homunkulů je **vývojový diagram** sestávající z řady bloků, z nichž každý specifikuje jednoho homunkula popisem jeho funkce, aniž by přitom jakkoli blíže určoval způsob, jakým je tato funkce realizována (tzv. **funkční abstrakce**). Kdybychom se však podívali na jednotlivé bloky vývojového diagramu pod větším rozlišením, zjistili bychom, že funkce každého z nich je realizována prostřednictvím další dělby práce přes další vývojové diagramy do stále menších a stále hloupějších homunkulů. Míru rozlišení takto můžeme zvyšovat až do chvíle, kdy se dobereme k homunkulům tak hloupým, že jediné co dokážou, je odpovědět ANO, nebo NE podle toho, zda obdržely, nebo neobdržely svůj signál (asi tak jako zámek dokáže „rozeznat“, zda do něj klíč pasuje, nebo nepasuje). Tito homunkulové jsou tak „omezení“, že jejich funkci může zastat i ten nejhloupější člověk nebo **stroj**. Správným sestavením armády takovýchto „hlupáků“ pak lze vytvořit systém, který jako celek plní funkci, která se vnějšímu pozorovateli jeví jako zřejmý projev inteligence (viz také obrázek 19b). Příkladem aplikace tohoto přístupu v teorii zrakového vnímání je **pandemonium Olivera Selfridge**, které sestává z několika vrstev démonů – jednoduchých procesů netrpělivě čekajících na svou příležitost započít svou jednoduchou a omezenou činnost, spočívající v signalizaci démonům v další vrstvě, že v kakofonii signálů (přicházejících z předchozí vrstvy démonů) rozpoznali zprávu, která je určena právě pro ně. Na obrázku 20 je znázorněno, jak takové pandemonium



Krok 1: Definice funkce, kterou má být systém schopen realizovat.



Krok 2: Rozložení funkce na řadu dílčích funkcí.



Krok 3: Opakování kroku 2, dokud nejsou jednotlivé dílčí funkce dostatečně jednoduché na to, aby je bylo možné realizovat pomocí nějakého technického zařízení.

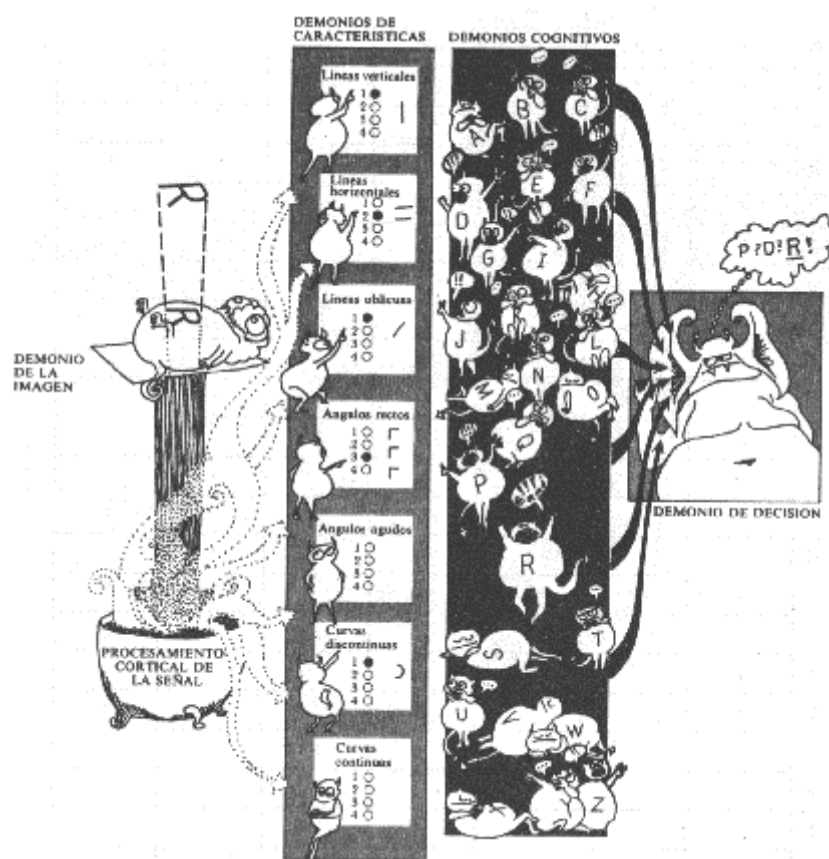
Obrázek 19b: Princip řešení problému s nekonečným regresem homunkulů, nazývaného také „Rylův regres“ podle filozofa Gilberta Ryla¹, který byl přesvědčen, že funkční analýza mentálních jevů – kdy jeden funkční pojem rozkládáme na soubor dalších funkčních pojmů – je odsouzena k nekonečnému regresu nevysvětlených funkčních pojmů. Řešení tohoto problému spočívá v tom, že funkční analýza se jednoduše zastaví na té úrovni, kde jsou jednotlivé funkce již dostatečně jednoduché na to, aby je bylo možné realizovat prostřednictvím nějakého technického zařízení. Metaforicky řečeno, myšlenkové zakázky pro pomyslného homunkula sídlícího v karteziánském divadle, kde se mají sbíhat veškeré informace z periferie, je potřeba rozdělit na řadu dílčích úkolů a ty pak rozdistribuovat („outsourcovat“) mezi větší počet (například neurálních) subdodavatelů. A jak v této souvislosti poznamenává Dennett (2004a, s. 238): „Jestliže se učiníte skutečně malými, můžete externalizovat téměř cokoli.“² (obrázek s úpravami převzat z <http://www.bcp.psych.ualberta.ca/~mike/>)

zpracovává písmeno „R“: Tzv. obrázkový démon přijímá senzorické signály ze sítnice a předává je do následující vrstvy tzv. znakových démonů specializovaných na detekci signálů indukujících přítomnost určitých geometrických prvků (svislá linie, vodorovná linie, šikmá

¹ Gilbert Ryle se proslavil rovněž svou sémantickou analýzou pojmu mysl, o kterém byl přesvědčen, že v základech jeho používání se skrývá tzv. *kategoriální chyba*, která má za následek zkreslený pohled na to, jaká je skutečná povaha lidské mysli. Podstatu kategoriální chyby Ryle ilustroval na následujícím příkladě: „Cizinci, který je poprvé v Oxfordu..., je předvedena spousta fakult, kolejí, knihoven, hřišť, muzeí, vědeckých pracovišť a úřadů. On se pak zeptá: „Ale kde je ta univerzita? Viděl jsem, kde bydlí studenti, kde pracuje tajemník, kde dělají vědci pokusy a všechno to ostatní. Ale ještě jsem neviděl tu univerzitu, ve které příslušníci vaší univerzity sídlí a pracují.““ Cizincova chyba spočívala v tom, že se domníval, že univerzita je nějaká další budova, na kterou si je možné šáhnout stejně jako na budovy fakult, knihoven atd. Univerzita je ale ve skutečnosti pouze souhrnný název pro všechny tyto budovy a instituce. Stejně chyby se podle Ryla dopouštíme tehdy, když o lidské mysli uvažujeme jako o jedné z mnoha komponent, ze kterých se člověk skládá. To pak vede k „představě osoby jako jakéhosi ducha záhadně ukrytého ve stroji“. Ve skutečnosti je mysl pouze souhrnné označení všeho toho, z čeho se člověk skládá a co dělá člověka člověkem, stejně jako univerzita představuje souhrnné označení pro koleje, laboratoře, hřiště atd. (všechny citace jsou z Ryleovy knihy *The Concept of Mind*, citováno podle Peregrin, 2005, s. 170, 173)

² „If you make yourself really small, you can externalize virtually anything.“

linie, pravý úhel, ostrý úhel atd.) v sítnicovém obraze. Tyto démoni pak předávají zprávu o své aktivaci do další vrstvy tzv. kognitivních démonů, kteří reagují pouze na určité specifické kombinace signálů z předchozí vrstvy démonů. Čím těsnější je korelace mezi přicházejícími signály a tou kombinací signálů, která daný kognitivní démon aktivuje, tím více je takový démon aktivní a tím je i silnější signál, který vysílá k tzv. rozhodovacímu démonovi, jehož rozhodnutí o tom, na jaké písmeno se to člověk vlastně dívá, je založeno na tom, který z kognitivních démonů v předchozí vrstvě vysílá nejsilnější signál. Obdobný způsob organizace (kognitivní) práce lze nalézt de facto ve všech tradičních i méně tradičních modelech kognitivních procesů – namátkou jmenujme například **produkční systém** Allena Newella a Herberta Simona (Simon, Newell, 1972), **teorii společenství myslí** Marvinina Minskyho (1985), **subsumpční architekturu** Rodneyho Brookse (1999) či **Copycat** Douglase Hofstadtera a Melanie Mitchellové (Mitchell, Hofstadter, 1990).



Obrázek 20: *Pandemonium sestávající z hierarchie obrázkových, znakových, kognitivních a rozhodovacích démonů.*

Aplikací tohoto principu dělby práce na procesy spojené s fungováním formálních systémů je možné delegovat procesy manipulace a transformace symbolických reprezentací na dobře zorganizovanou armádu hloupých strojů, a celý formální systém tak **zautomatizovat**. Takový automatický formální sy-

stém je pak stroj či zařízení, jehož některé součástky nebo fyzikální stavy, ve kterých se může nacházet, představují znaky vybraného formálního systému, a jehož chod zajišťuje automatickou manipulaci a transformaci těchto znaků v souladu s pravidly daného formálního systému. Haugeland (1997) automatický formální systém přirovnává k šachovnici plné figurek, které se pohybují zcela samy od sebe, bez jakékoli intervence ze strany hráčů, a

přítom vždy v souladu s pravidly šachu; nebo také k „magické“ tužce, která dokáže na kusu papíru sama vyvozovat důkazy a teoremy v rámci určitého formálního systému bez jakékoli pomoci lidského matematika či logika.

Podle Dennetta (2004b) byl prvním pokusem o formulaci a popis takového mechanismu, který by umožňoval automatickou manipulaci a transformaci mentálních reprezentací, **asocianismus**. Podle jeho představitelů - mezi které patřil například skotský filozof *David Hume* - podstata lidského myšlení spočívá ve schopnosti různým způsobem spojovat a kombinovat do složitějších idejí jednotlivé dojmy (počítky a vjemy) a ideje (kopie dojmů), které tvoří základní stavební materiál lidského duševního života. Aby nebyl tento proces závislý na inteligenci nějakého centrálního homunkula, bylo třeba vymyslet mechanismus, který by zaručil, že se dojmy a ideje budou spojovat a kombinovat samy, bez jakékoli cizí pomoci a asistence. Tímto mechanismem byly pro asocianisty tzv. **principy sdružování idejí**, neboli **asociační zákony**, z nichž některé formuloval již *Aristoteles* ve 4. století př.n.l. (například **princip podobnosti** a **princip kontiguitu**, tj. styčnosti v čase a prostoru) (Störig, 1993). Tyto principy fungují jako určitý specifický typ vazeb v „mentální chemii“, jejíž zákonitosti jsou zodpovědné za nenáhodný způsob, jakým se ve vědomé mysli člověka sdružují a kombinují jednotlivé dílčí dojmy a ideje. Jakkoli jsou tyto návrhy asocianistů – alespoň tedy ve své původní podobě - formulovány dosti vágně (takže je možné je jen velice obtížně podrobit rigoróznějšímu vědeckému zkoumání), představují správný směr uvažování, který nám umožňuje se zbavit strašáka v podobě nekonečného regresu všemocných a vševědoucích homunkulů. Myšlenky asocianistů byly později precizovány v rámci myšlenkové linie **reflexologie-behaviorismus-konekcionismus**, která na asocianismus volně navazovala: Myšlenkami asocianistů se přímo inspiroval ruský fyziolog *I. P. Pavlov* při svých pokusech s jednoduchým podmiňováním, které zase byly inspirací pro americké behavioristy, kteří studovali složitější schémata podmiňování. Někteří behavioristé se později pokusili propojit poznatky behaviorismu s poznatky **neurovědy** – tehdy se nově rodící vědecké disciplíny o lidském mozku. Takto například *Donald Hebb* ukázal, že mechanismus jednoduchého podmiňování je možné implementovat prostřednictvím upravování vah spoju (synapsí) mezi neurony. Toto tzv. *hebbovské pravidlo učení* – spolu s dalšími učícími algoritmy – se později stalo důležitou součástí konekcionistických (výpočetních) modelů některých kognitivních procesů.¹

¹ Nové způsoby uchopení některých myšlenek asocianistů se otevřely také s rozvojem nástrojů tzv. **soft-computingu**, které umožňují vytvářet přesné výpočetní modely jevů, které jsou inherentně neurčitě, přibližné a „náchylné“ k jen částečné pravdivosti. U nás například psycholog *Petr Šlechta* (1998) ve své

1.2.2.1 Turingův stroj

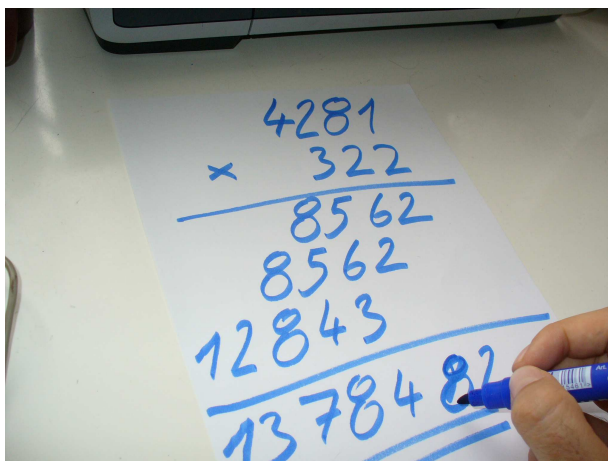
Jiným a z hlediska vzniku výpočetní teorie myslí mnohem důležitějším (a úspěšnějším) pokusem o konstrukci automatického formálního systému je tzv. **Turingův stroj**, teoretický předobraz dnešních počítačů. Turingův stroj není klasické zařízení sestavené z mechanických součástí; jedná se o čistě teoretické zařízení (je to de facto svého druhu myšlenkový experiment), které britský matematik Alan Turing navrhl v rámci svých snah formalizovat pojem výpočtu a přispět tak ke zkoumání tzv. problému *vypočitatelnosti*, tedy otázky, co je a co není možné vypočítat (Hillis, 2003).

Turingův stroj je výsledkem Turingovi snahy nalézt odpověď na otázku „*Co to vlastně dělám, když počítám?*“. Při návrhu svého imaginárního stroje vycházel Turing z introspektivní analýzy posloupnosti svých myšlenkových pochodů během provádění numerických výpočtů a řešení různých matematických problémů. Turing zjistil, že – bez ohledu na to, jak komplikovaný je výpočet, který provádí - při počítání nedělá vlastně nic jiného, než že zjišťuje, jaké transformační pravidlo má právě použít, poté vybrané pravidlo aplikuje, zapíše výsledek, podívá se na něj a zjišťuje, jaké transformační pravidlo by měl na tento výsledek použít atd. dokud nedospěje k řešení celého problému (Dennett, 1993). Sebekomplikovanější výpočet tak lze podle Turinga zredukovat na sadu jednoduchých instrukcí stigmergického algoritmu, který k řízení a koordinaci posloupnosti činností využívá jednoduchá „lokální“ pravidla pro tvorbu a modifikaci různých vnějších struktur, kterými jsou v případě výpočtu různé symbolické reprezentace (napsané například na kusu papíru), ve stylu:

1. *Vytvoř, například napiš na papír, vzorec či řetězec symbolů.*
2. *Prohlédni si napsané symboly.*
3. *Identifikuj ty symboly, které jsou spojeny s pravidly pro tvorbu dalších symbolů.*
4. *Napiš další symboly.*
5. *Vrať se ke kroku č. 2.*

Jinými slovy, „*Turing si uvědomil, že každý výpočet (či obecněji každé řešení) začíná nějakými vstupními daty, které si můžeme představit znak po znaku zapsané na papírové pásce, a končí nějakým výsledkem, který si opět můžeme představit v této podobě. Výpočet je*

diplomové práci demonstroval, jak lze některé asocianistické principy sdružování idejí realizovat v rámci tzv. **fuzzy množinového kognitivního (expertního) systému**.

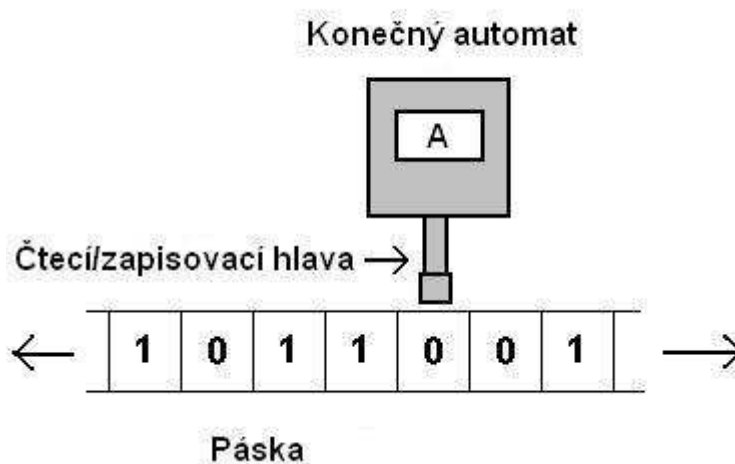


Obrázek 21: Procedura dlouhého násobení, při které člověk využívá tužku a papír k zápisu mezivýsledků jednotlivých výpočetních kroků, je typickým příkladem užití tzv. stigmergického algoritmu, který k organizaci komplexní posloupnosti činností využívá jednoduchá „lokální“ pravidla pro tvorbu a modifikaci různých vnějších (zde symbolických) struktur.

tedy z tohoto pohledu přechod od jedné sekvence znaků na pásce k jiné; a Turing usoudil, že ať už ten přechod provádíme jakkoli, na té nejelementárnější úrovni se nemůže než skládat z několika operací toho typu, jako je přečtení nějakého existujícího symbolu, posun pásky o jednu pozici tam či zpátky a zapsání nového symbolu či přepsání starého.“ (Peregrin, 2003) Produktem Turingovi „destilace“ podstaty matematického myšlení z proudu lidského vědomí je tak soubor elementárních mentálních operací, které dostačují k realizaci jakéhokoli výpočtu a které jsou přitom tak jednoduché, že je zvládne i stroj.

Turing tento svůj náhled formuloval v podobě specifikace vlastností imaginárního výpočetního zařízení, které je zjednodušenou a zidealizovanou verzí hypotetického hyperracionálního matematika provádějícího své výpočty. Tomuto imaginárnímu zařízení se dnes říká Turingův stroj (dále TS). TS se skládá z nekonečně dlouhé pásky, konečného automatu a ze čtecí/zapisovací hlavy. Páska má funkci paměti; slouží tedy jako úložiště pro data, která jsou na pásce reprezentována prostřednictvím určitého pevného souboru arbitrárních symbolů. Čtecí/zapisovací hlava dokáže z daného místa pásky přečíst symbol, dokáže symbol na pásku také zapsat nebo ho z pásky naopak vymazat; čtecí/zapisovací hlava se také dokáže pohybovat, vždy maximálně o jedno políčko pásky doprava nebo doleva. Konečný automat funguje jako jednoduchý procesor: Konečný automat (dále KA) je obecně zařízení, jehož výstup (ať už je to cokoli) nezávisí pouze na aktuálním vstupu, ale také na dřívější historii vstupů, která tvoří stav KA; KA má paměť, která mu umožňuje sledovat, v jakém stavu se právě nachází; KA je konečný proto, že má vždy omezený/konečný počet

stavů, ve kterých se může nacházet; vedle toho vždy existuje soubor povolených vstupů, které mění stav KA, a soubor možných výstupů, který je dán aktuálním stavem KA a aktuálním vstupem do KA. Jednoduchým příkladem takového KA je kuličkové pero, které se může nacházet pouze ve dvou možných stavech (vysunutá kulička/zasunutá kulička pera), které se mění zmáčknutím tlačítka (vstup); výstupem kuličkového pera je buď vysunutí, nebo zasunutí kuličky pera; díky své konstrukci si kuličkové pero „pamatuje“ jaká je historie jeho vstupů (tj. zda byl počet zmáčknutí tlačítka sudý nebo lichý), resp. to, v jakém stavu se právě nachází (Hillis, 2003). V případě Turingova stroje KA dokáže na základě znalosti svého aktuálního stavu a právě přečteného symbolu určit, jaký (a zda vůbec) symbol se má zapsat na pásku, jak



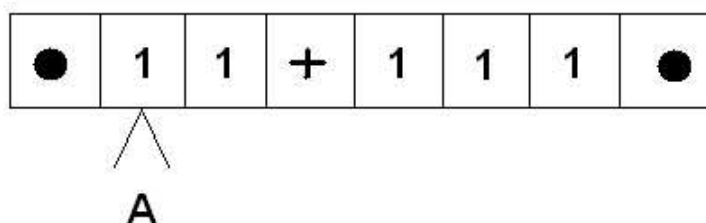
Obrázek 22: Turingův stroj.

(a zda vůbec) se má změnit stav KA a kam (a zda vůbec) se má posunout čtecí/zapisovací hlava. K úplné specifikaci TS jako stroje vykonávajícího určitý výpočet je ještě potřeba definovat pevný soubor symbolů, kterými se bude kódovat vstup a výstup KA (na obrázku 22 to jsou číslice 1 a 0), sadu možných stavů KA (například A a B) a přechodovou funkci (sadu základních instrukcí), kterou se řídí operace KA při transformaci vstupů (aktuální stav KA-aktuálně načtený symbol) na výstupy (zápis symbolu na pásku-změna stavu KA-přesun čtecí/zapisovací hlavy). Řekněme, že bychom chtěli takový TS přimět počítat dvojice celých kladných čísel¹. Tento typ výpočetní úlohy je možné TS prezentovat na paměťové pásce v podobě posloupnosti symbolů „•“, „1“ a „+“, které reprezentují nejen zadaný problém, ale i návod na jeho řešení. Tímto způsobem zakódovaná početní úloha $2 + 3$ bude na pásce TS vypadat tak, jak je to zachyceno na obrázku 23a. Na obrázku 23a jsou ještě další dva znaky, které označují aktuální polohu čtecí/zapisovací hlavy („^“) a jeden ze dvou možných

¹ Uvedený příklad je mírně upravenou verzí TS, který ve své knize uvádí filozof J. Kim (1996, s. 80-85).

aktuálních stavů KA („A“). Aby TS dokázal tímto způsobem reprezentovanou úlohu vyřešit, potřebuje ještě tu správnou sadu instrukcí definujících konečný soubor primitivních operací, které by mu umožnily na základě dat z pásky s jednotlivými symboly různě manipulovat a transformovat je na jiné symboly způsobem odrážejícím některé matematické zákonitosti. Tyto instrukce by mohly vypadat následujícím způsobem:

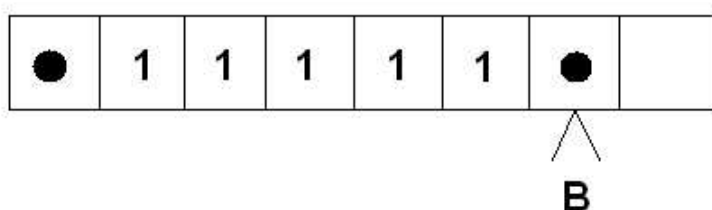
- *Instrukce č. 1:* Jestliže se KA nachází ve stavu A a načteným symbolem je „1“, potom ať se čtecí/zapisovací hlava posune o jednu pozici doprava a KA ať zůstane ve stavu A.
- *Instrukce č. 2:* Jestliže se KA nachází ve stavu A a načteným symbolem je „+“, potom ať čtecí/zapisovací hlava přepíše tento symbol na „1“, posune se o jednu pozici doprava a KA ať zůstane ve stavu A.
- *Instrukce č. 3:* Jestliže se KA nachází ve stavu A a načteným symbolem je „•“, potom ať se čtecí/zapisovací hlava posune o jednu pozici doleva a stav KA ať se změní na stav B.
- *Instrukce č. 4:* Jestliže se KA nachází ve stavu B a načteným symbolem je „1“, potom ať čtecí/zapisovací hlava tento symbol vymaže, nahradí ho symbolem „•“ a ukončí veškeré další změny.



Obrázek 23a (s mírnými úpravami převzato z Kim, 1996)

Celý výpočet probíhá postupně, krok za krokem, **sériově** - vždy se načítá právě jeden symbol a vždy se provádí právě jedna operace. Při řešení úlohy TS začíná u symbolu „1“, který je na levém okraji pásky. Poté prohledává pásku směrem doprava, dokud nenarazí na symbol „+“, který přepíše symbolem „1“. TS potom pokračuje v prohledávání pásky směrem doprava, aby našel symbol „•“, který označuje konec druhého sčítance. Jakmile na něj narazí, vymaže ho, posune se o jedno políčko doleva a přepíše symbol „1“ na symbol „•“. V této chvíli (viz obrázek 23b) je na pásce zakódována odpověď ve stejném formátu, v jakém byla zakódována

početní úloha.¹ Tento jednoduchý příklad názorně ilustruje, jak soubor jednoduchých formálních operací, které pracují čistě s formálními/syntaktickými vlastnostmi symbolických struktur, dokáže podporovat a zachovávat důležité sémantické zákonitosti vybrané problémové domény. V případě TS navíc tento proces probíhá zcela automaticky a bez jakéhokoli zásahu zvenčí.



Obrázek 23b (s mírnými úpravami převzato z Kim, 1996)

Obdobným způsobem jako sčítací TS je možné si nadefinovat Turingovy stroje, které by dokázaly odčítat, dělit, násobit nebo také hrát šachy, dělat lékařské diagnózy, hledat nová naleziště nerostných surovin, odhalovat rizikové klienty bank nebo řešit jiné, dostatečně dobře specifikované problémy; zde je důležité si uvědomit, že TS (a tím pádem také počítač, který je jeho „hmotnou inkarnací“) není nějaký numerický kalkulátor - přestože numerické výpočty také dokáže provádět -, ale především **informační** či **symbolový procesor** (*information/symbol processing system*), neboť to, s čím TS na té nejnižší úrovni operuje/počítá/manipuluje, nejsou čísla, ale informace, resp. arbitrární symboly, jejichž různé kombinace mohou (pod tou správnou interpretací) znamenat cokoli – mohou to být čísla, ale také například pozice šachových figurek na šachovnici, symptomy nemoci, názvy lékařských diagnóz, možné způsoby léčby, vlastnosti půdy nebo bonita bankovního klienta.

Nejdůležitějším Turingovým objevem bylo, že stejně jako si můžeme nadefinovat TS, který dokáže sčítat nebo násobit, můžeme si také nadefinovat jeden zvláštní druh TS, tzv. **Univerzální Turingův stroj** (dále UTS), který má tu správnou sadu základních instrukcí/operací, aby dokázal na základě symbolického popisu chování jiného TS (zaznamenaného na pásce) napodobit vstupně-výstupní funkci, kterou tento jiný TS počítá. A protože UTS takto dokáže napodobit chování jakéhokoli TS, znamená to, že **UTS dokáže vyřešit jakýkoli dostatečně dobře specifikovaný výpočet/problém** (tzv. *Church-Turingova teze*). „Turing [takto] ukázal, že jakýkoli... problém lze řešit zapsáním znaků na pásku

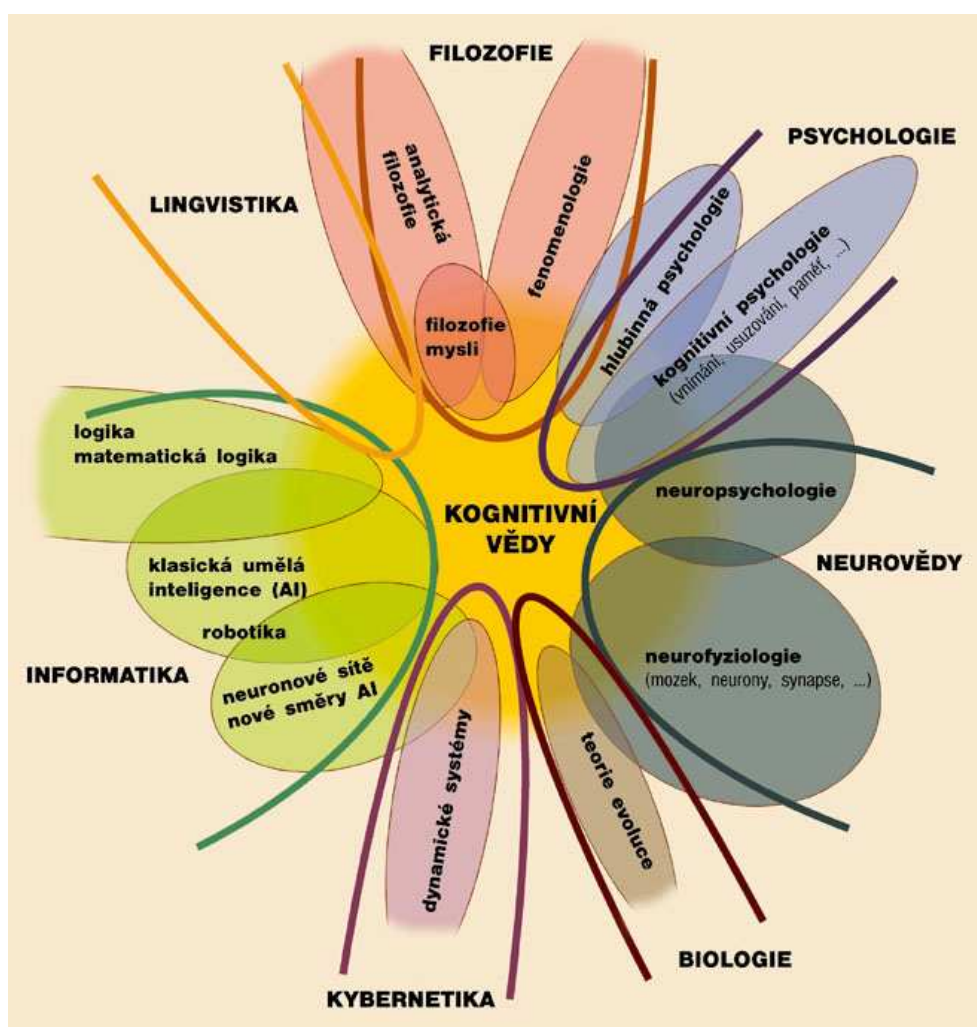
¹ Podobný TS, který dokáže sčítat celá čísla, lze nalézt také na příloženém CD v podobě jednoduché excelovské aplikace pod názvem *Turingův_stroj*.

Turingova stroje – znaků, které nespecifikují jen problém, ale také postup, jak ho řešit. Turingův stroj spočítá odpověď pohyby tam a zpět po pásce, čtením a zapisováním znaků, dokud není řešení zapsané na pásce.“ (Hillis, 2003, s. 70)

Z Turingova objevu tedy vyplývá, že může existovat stroj, jehož normální činnost a fyzikální vlastnosti implementují syntaktické/formální operace definované na souboru arbitrárních symbolických struktur, a že takový stroj dokáže vyřešit jakýkoli dostatečně dobře specifikovaný problém. To, co je na takovém stroji přitom důležité, není to, z čeho je postaven, ale pouze jeho formální/syntaktická struktura či organizace, neboť pouze na ní závisí, zda stroj uspěje, nebo neuspěje při zachování sémantických charakteristik vybrané problémové domény: Podobně jako se nic nezmění na podstatě šachu když místo dřevěných figurek na dřevěné šachovnici budeme hrát s mezigalaktickými koráby a 64 galaxiemi, tak se nezmění ani podstata výpočtu, který TS provádí, jestliže TS bude postaven z různých materiálů. To, na čem záleží je pouze abstraktní síť povolených tahů a přechodů mezi jednotlivých stavy daného systému, který může být sestaven z mnoha různých materiálů. Tímto vznikla „*nová úroveň analýzy, která je nezávislá na fyzice, ale přesto je ve své podstatě mechanistická... je to věda o struktuře a funkci oddělené od materiální substance*“¹ (Pylyshyn, 1986, s. 68). A právě tato nová úroveň analýzy se na přelomu 50. a 60. let 20. století stala hlavním předmětem nově vznikající kognitivní vědy, která se skládá z řady dílčích disciplín (viz obrázek 24), z nichž většina více či méně sdílí předpoklad, že lidská mysl je formou výpočtu - ve smyslu manipulace a transformace syntaktických (symbolických/reprezentačních) struktur – a že zařízením, které tento výpočet implementuje a provádí, je lidský mozek, jehož kauzální struktura (tzn. fyzikální stavy nervové tkáně a jejich změny v čase) zrcadlí specifickou abstraktní/formální strukturu výpočtu (jednotlivé stavy/prvky/symbole a přechody mezi nimi), který je lidským myšlením. Lidský mozek je tak tedy počítačem ve stejném smyslu jako elektronická kalkulačka (viz s. 5): U obou zařízení existuje spolehlivý a systematický způsob interpretace jejich různých fyzikálních stavů (značky a číslice na displeji a klávesnici, resp. vzorce neuronální aktivity) jako něčeho, co reprezentuje nějaké jiné entity (čísla, resp. myšlenky a přesvědčení), a obě zařízení jsou sestrojena způsobem, který jim umožňuje – při odpovídající interpretaci – zrcadlit smysluplné (sémantické) zákonitosti v matematické, resp. mentální oblasti. Jinými slovy, obě zařízení obsahují určité vnitřní fyzikální stavy, které reprezentují předměty a události vnějšího světa, a obě využívají různých mechanických procedur k realizaci přechodů mezi těmito stavy, tak

¹ „...a new level of analysis, independent of physics yet mechanistic in spirit... a science of structure and function divorced from material substance.“

aby mohly takto reprezentované informace racionálně využívat. Taková chytrá kalkulačka pak dokáže na výstupu vyprodukovat „8“, jestliže na vstupu obdrží „5 + 3 =“, a inteligentní mozek zase bude vědět, že není příliš chytré vstupovat do jeskyně, do které před chvílí viděl vstupovat tři medvědy, a nedlouho poté dva medvědy tuto jeskyni opouštět.¹ To, o co se přitom kognitivní věda - alespoň ve své tradiční podobě - nejvíce zajímá, je abstraktní specifikace výpočtu (symbolické reprezentace a transformační algoritmy, které se na ně aplikují), který toto inteligentní chování umožňuje; již mnohem méně se pak zajímá o konkrétní způsob fyzické implementace takového výpočtu – neboť jestliže výpočet můžeme realizovat více různými způsoby, potom to, co je důležité, je pouze abstraktní forma či organizace výpočtu, ne jeho konkrétní fyzická implementace.



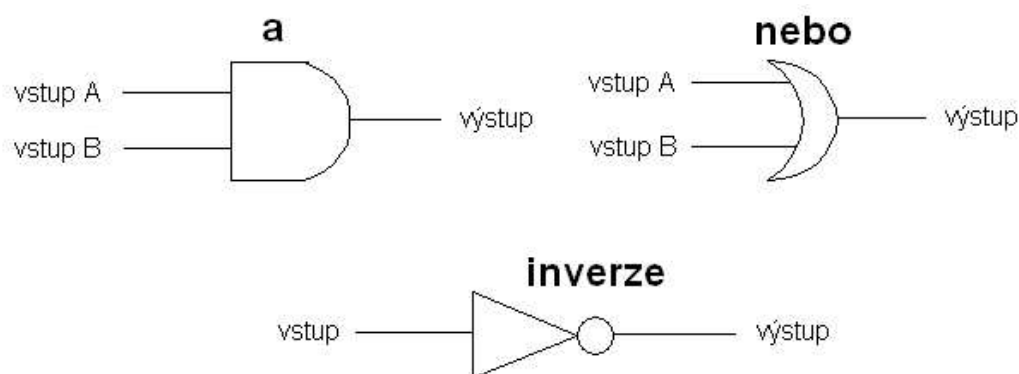
Obrázek 24: Struktura kognitivní vědy. (Havel, 2000)

¹ Neschopnosti některých ptáků takového jednoduchého numerického usuzování využívají často ornitologové, kteří se díky tomu pro ptáky stávají „neviditelnými“: „Dva z vás jdou do úkrytu a pouze jeden z vás ho opustí. Bez tohoto triku by byli ptáci ve střehu kvůli schovanému, protože „vědí“, že někdo tam vešel. Když ale jedna osoba odejde, „usoudí“, že odešly obě.“ (Dawkins, 1996, s. 70)

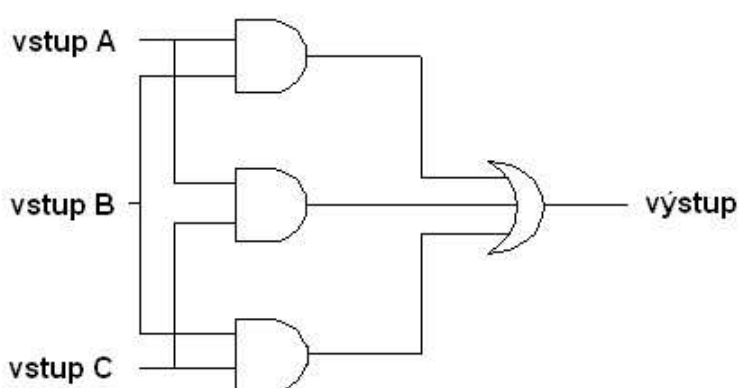
1.2.2.2 Počítače, von Neumannovská architektura a virtuální stroje

Posledním krokem při formování kognitivní vědy bylo sestrojení skutečného programovatelného digitálního počítače, který by kognitivním vědcům umožnil vytvářet modely různých mentálních procesů. Konstrukce tohoto počítače byla založena na myšlenkách, které Alan Turing vtělil do svého imaginárního zařízení. Překlopení tohoto abstraktního myšlenkového experimentu ve skutečný stroj umožnili zejména *John von Neumann*, který Turingovy myšlenky převedl v (stále ještě dost abstraktní) návrh konkrétní výpočetní architektury, a *Claude Shannon*, který ukázal, jak fyzicky realizovat prvky takové výpočetní architektury.

V roce 1940 Claude Shannon, zakladatel oboru teorie informace, ve své práci *A Symbolic Analysis of Relay Switching Circuits* (Symbolická analýza obvodů s přepínáním pomocí relé) prokázal, že lze využít jednoduché elektrické obvody (spínače a vodiče) k implementaci výroků v booleovské algebře (viz s. 44-46): Pravdivostní hodnoty jednotlivých proměnných v logických výrocích (resp. bity - základní jednotky informace) jsou zde reprezentovány elektrickým proudem, který obvodem buď prochází, nebo neprochází; uzavřené spínače (procházející proud) odpovídají pravdivým (1) a otevřené spínače (neprocházející proud) nepravdivým (0) proměnným booleovské algebry. Výroky v booleovské algebře je pak možné převést do specifického uspořádání spínačů, které mohou být zapojeny buď sériově, nebo paralelně. Při sériovém zapojení dvou spínačů proud obvodem protéká pouze tehdy, když jsou oba spínače uzavřeny. Takové zapojení implementuje logickou funkci „a“. Při paralelním zapojení spínačů proud obvodem prochází kdykoli je jeden nebo oba spínače uzavřeny. Tento způsob uspořádání spínačů odpovídá logické funkci „nebo“. Podobně jednoduchým způsobem je možné implementovat rovněž logickou funkci „negace“. Takto zapojené spínače tvoří tři základní logické jednotky (tzv. **logická hradla**) „a“, „nebo“ a „invertor“, jejichž kombinací lze vytvořit elektrické obvody reprezentující další logická pravidla a funkce. **Fakticky lze tímto způsobem implementovat jakoukoli binární logickou funkci** (tj. funkci určenou vyhledávací tabulkou tvořenou binárním symbolickým kódem), **kteřou je možné zapsat jako přesný logický výrok** (Hillis, 2003). Na obrázku 26 je takto schematicky zachycen jednoduchý elektrický obvod sestavený z několika logických hradel, která společně implementují hlasovací funkci, kterou zachycuje výrok $(A \cdot B) + (A \cdot C) + (B \cdot C)$ a jejíž vstupně/výstupní (také pravdivostní nebo vyhledávací) tabulku je možné nalézt na straně 46. Kombinací bitů (dvou symbolů 1/0) lze reprezentovat libovolná data a informace – čísla, obrazy, zvuk, tahy na šachovnici, gramatické kategorie atd. Počet bitů nutných ke



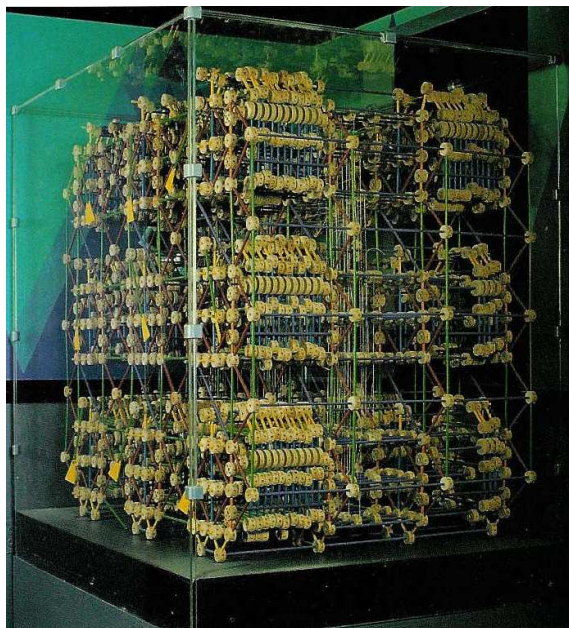
Obrázek 25: Logická hradla. Schematické znázornění elektrických obvodů implementujících logické jednotky „a“, „nebo“ a „inverze“. (převzato z Hillis, 2003)



Obrázek 26: Implementace hlasovací funkce prostřednictvím logických hradel „a“ a „nebo“. (převzato z Hillis, 2003)

kódování dat přitom závisí na počtu zpráv či sdělení, která je potřeba rozlišit. Například binární funkce, která by měla za úkol sčítat dvojice čísel z množiny čísel 0-99, by musela používat minimálně osmibitové kódování, které je schopno rozlišit 2^8 (tj. 256) možností, tak aby bylo možné zakódovat i nejvyšší možný součet dvou čísel (198). Všechna data, která jsou reprezentována tímto způsobem, tedy prostřednictvím **binárního kódu**, mohou být zpracována užitím nějaké binární logické funkce, kterou lze vždy implementovat určitou kombinací základních logických jednotek „a“, „nebo“ a „inverze“. Tyto logické jednotky tvoří základní funkční jednotky všech digitálních počítačů: Na té nejnižší úrovni nejsou digitální počítače ničím jiným než souborem **vodičů** a **spínačů**, které vedou a řídí určitou **mediační látku**, která slouží k posílání jedné ze dvou možných zpráv (1/0, Pravda/Nepravda) od jednoho spínače ke druhému. V dnešních počítačích se (z důvodu urychlení výpočetních procesů) jako mediační látka používá elektřina; ta se počítačem šíří po obvyklých elektrických vodičích a její tok je regulován křemíkovými tranzistory (v minulosti to byly např. elektromechanická relátka nebo elektronky). Správná funkce počítače je však na této

konkrétní implementační technologii zcela nezávislá – důležitá je forma, nikoli konkrétní způsob její realizace. Takto lze například sestavit počítač, který by kódoval bity nikoli prostřednictvím elektrického napětí, ale pomocí vodního tlaku; spínače a vodiče takového počítače by pak sestávaly z hydraulických ventilů a vodovodního potrubí. K reprezentaci bitů by také bylo možné využít mechanického tlaku, převáděného a regulovaného systémem tyček a pružin (viz obrázek 27). Paralelním nebo sériovým zapojením spínačů se implementují booleovské logické funkce „a“, „nebo“ a „inverze“. Z těchto základních logických jednotek



(logických hradel) jsou pak sestaveny dvě hlavní komponenty programovatelného digitálního počítače – **procesor** (konečný automat) a **paměť** (resp. registry, ze kterých se paměť skládá) (Hillis, 2003; White, 1994).

Obrázek 27: Počítač sestavený z tyček a pružinek stavebnice Tinkertoy. Počítačový vědec Daniel W. Hillis postavil s pomocí několika dětských stavebnic Tinkertoy mechanický počítač, který dokáže úspěšně hrát piškvorky (jejich jednodušší americkou verzi, která se hraje na mřížce 3x3). Činnost tohoto počítače je – stejně jako činnost jakéhokoli jiného počítače - založena na

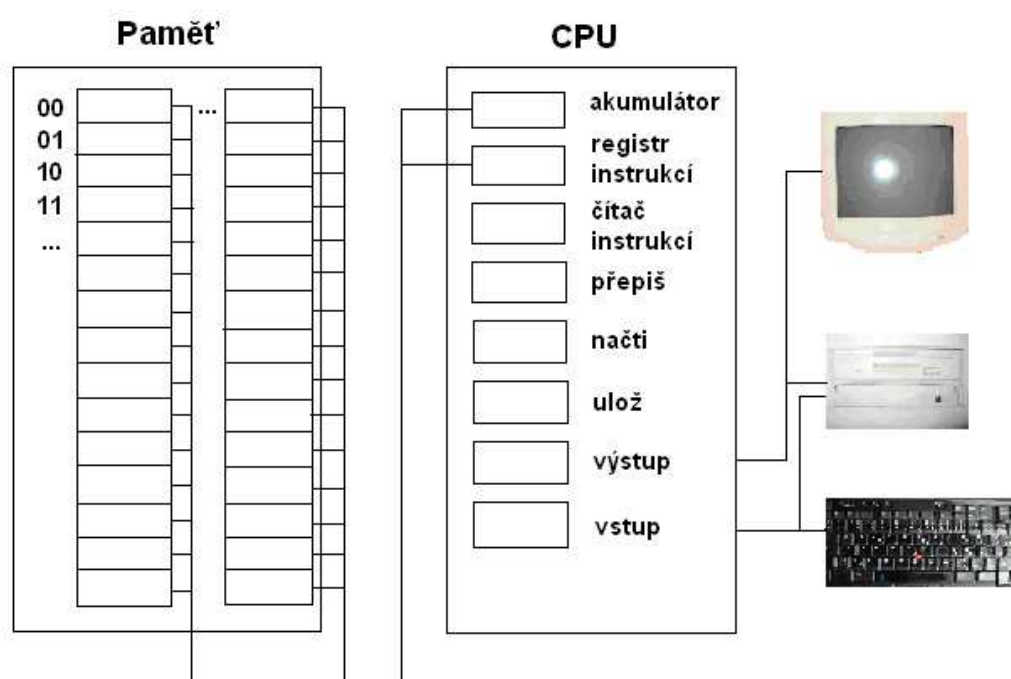
kombinaci velkého množství základních logických jednotek, které jsou zde implementovány prostřednictvím mechanického tlaku, kterým na sebe jednotlivé tyčky navzájem působí.

Valná většina současných i minulých počítačů má nebo měla tzv. von Neumannovskou výpočetní architekturu, která je přímo inspirována výpočetní architekturou Turingova stroje. Ten sestává z **konečného automatu** a z **paměti**, která je k němu připojena. TS počítá tím způsobem, že konečný automat **sériově** (postupně, jednu za druhou) provádí některou z několika jednoduchých **operací** na datech (**data** v úzkém slova smyslu) a na základě dat (**instrukce**), která jsou uložena v paměti, ze které konečný automat data postupně načítá nebo je v ní nějak mění. Počítač s von Neumannovskou architekturou (bez periferních zařízení jako je klávesnice, obrazovka nebo tiskárna) má v podstatě stejnou strukturu – tzn. že se skládá z konečného automatu (procesoru nebo také CPU¹) a k němu připojené paměti; toto **oddělení procesoru od paměti** je charakteristickým rysem všech počítačů s von Neumannovskou architekturou. Rovněž způsob, jakým počítač s von Neumannovskou architekturou počítá je de facto totožný s tím, jak počítá TS – CPU sériově provádí některou ze základních operací na

¹ Central Processing Unit

datech, která (rovněž sériově) načítá z paměti, ve které jsou uložena; rozhodování o tom, kterou operaci provést, kdy a na čem, se děje na základě instrukcí, které jsou – stejně jako zpracovávaná data - uložena v paměti a z ní také načítána.

Počítač s von Neumannovskou architekturou je schematicky zachycen na obrázku 28: Nalevo je paměť, která se skládá z registrů. V registrech jsou v podobě posloupnosti bitů (tzv. *slov*, která mají v současných počítačích většinou 32 bitů) zakódována data, která se mají zpracovat, a instrukce, které CPU říká, jakou operaci má vykonat. Každý registr navíc obsahuje svojí vlastní adresu - posloupnost bitů, která daný registr identifikuje (lokalizuje) a zpřístupňuje ho CPU, která tak může jeho obsah načíst nebo do něj naopak něco zapsat. Data lze přitom vždy načítat pouze z jednoho registru najednou; to samé platí i pro zapisování do



Obrázek 28: von Neumannovská architektura. (s úpravami převzato z Dennett, 1993, s. 213)

paměťového registru. Napravo se nachází CPU; ta se skládá ze tří registrů, které se nazývají *akumulátor*, *registr instrukcí* a *čítač instrukcí*, a z logických jednotek, které implementují určitý konečný soubor základních operací, jejichž kombinací (která se řídí instrukcemi načítanými z paměti) je možné realizovat jakýkoli výpočet. Počítání takového počítače probíhá tak, že z paměťových registrů se načítají data do akumulátoru a instrukce do registru a čítače instrukcí. Tyto tři registry, ve kterých se může najednou objevit právě jen jedno slovo (posloupnost bitů z paměťového registru), jsou příčinou známého **von Neumannského úzkého hrdla**, kterým je nucena procházet veškerá výpočetní aktivita počítače. Ovšem pro

běžného uživatele je toto omezení - díky rychlosti implementační technologie, která umožňuje provádět miliony základních operací za sekundu - prakticky neviditelné.¹ Když se instrukce elektronicky překopíruje do registru instrukcí, specializované logické jednotky tuto instrukci provedou. Například jestliže instrukce říká (přeloženo do češtiny) „Vymaž obsah akumulátoru!“, potom se v CPU aktivují logické jednotky, které nastaví všechny bity (jejichž počet závisí na tom, o kolika-bitový počítač se jedná) v akumulátoru na 0. Když instrukce bude „Přičti obsah registru s adresou 10 k obsahu akumulátoru“, potom jiné specializované logické obvody načtou obsah tohoto registru a přičtou ho k obsahu akumulátoru. Čítač instrukcí pak vždy obsahuje adresu paměťového registru, ve kterém je uložena další instrukce, kterou CPU načte a následně i vykoná. Mezi základní operace, které CPU provádí, patří většinou aritmetické operace (sčítání, odčítání, dělení a násobení), operace různě manipulující s daty (načítání, ukládání, mazání...) a podmínkové instrukce typu „Jestliže má akumulátor ten či onen obsah, potom načti instrukci z registru s adresou 11, jinak načti instrukci z registru s adresou 01“ (Dennett, 1993). Počet těchto základních operací se může u jednotlivých počítačů lišit; může jich být pouze 16 (počítače s tzv. redukováným souborem instrukcí, tzv. RISC²) nebo také několik stovek (počítače s kompletním souborem instrukcí, tzv. CISC³). Velikost instrukčního souboru ale není pro výkon a schopnosti počítače nijak důležitá, neboť s dostatečně velkou instrukční sadou lze simulovat jakoukoli jinou instrukční sadu (Hillis, 2003). Každá ze základních operací je reprezentována v tzv. **strojovém kódu** – základním programovacím jazyce. V tomto jazyce je každé primitivní operaci přiřazen určitý binární kód (sada jedniček a nul), který - když se z paměťového registru načte do registru instrukcí v CPU - aktivuje specializované logické obvody implementující výše uvedené základní operace. Programátoři, kteří specifikují/programují práci, kterou má počítač vykonat, většinou nepracují na tak nízké úrovni, kterou představuje strojový kód počítače. Obvykle pracují s více či méně specializovanými vysokoúrovňovými programovacími jazyky (jako jsou např. BASIC, FORTRAN, C++, LISP nebo Java), jejichž základní výrazy (primitiva) jsou definovány jako posloupnosti primitivních operací zakódovaných ve strojovém kódu. Většina těchto programovacích jazyků je rozšiřitelná, tzn. že v nich lze definovat nová slova kombinací primitiv anebo jiných slov již nějak definovaných. Tyto **vysokoúrovňové programovací jazyky** tak představují uživatelsky vstřícné rozhraní, které programátorovi nabízí speciální jazyk (sestavující z určité slovní zásoby a syntaxe), kterým může počítači

¹ Rychlost současných počítačů je tak obrovská, že výsledné chování počítače může činit dojem, že počítač dělá více věcí najednou a nikoli jednu po druhé.

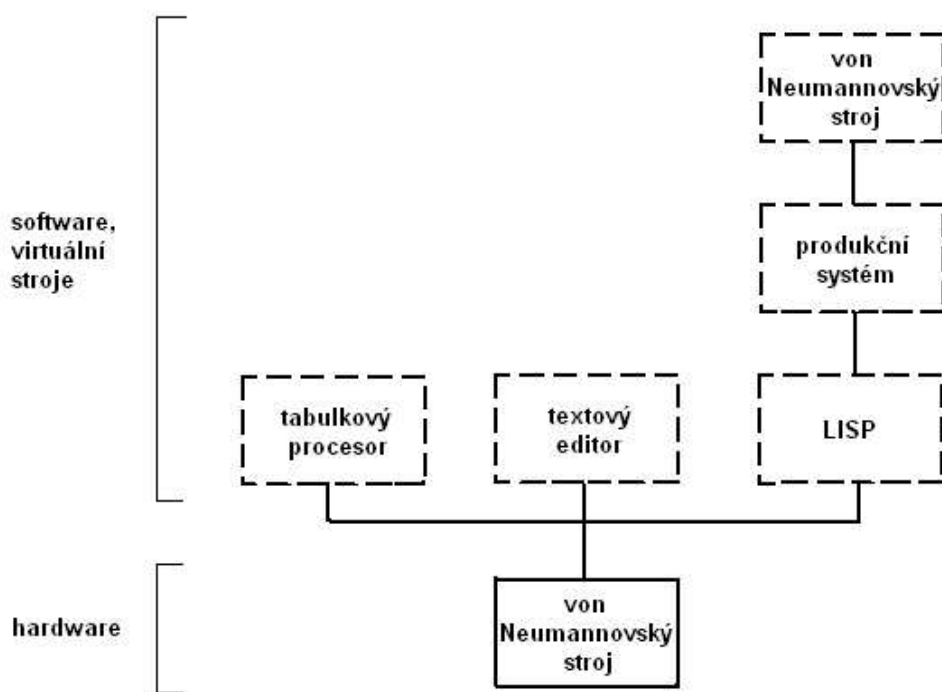
² *Reduced Instruction Set Computer*

³ *Complete Instruction Set Computer*

relativně pohodlně sdělovat, jakou práci po něm chce provést. Aby počítač napsaný program vykonal, je potřeba ho přeložit zpátky do strojových instrukcí, což se děje pomocí speciálního programu, kterému se říká buď *překladač*, nebo *interpret* - to podle toho, zda se převod provádí ještě před spuštěním samotného programu, nebo až po něm (Hillis, 2003).

Všechny digitální počítače, se kterými se člověk může běžně setkat, sdílí tuto základní výpočetní architekturu. Klíčovou vlastností této výpočetní architektury je její ohromná **flexibilita**. Přestože v sobě obsahuje řadu pevných, neměnných a natvrdo-zadrátovaných prvků, díky paměti, ve které mohou být uložena jak data tak i programy, ji lze snadno přeměnit na různá výpočetní zařízení s různou výpočetní architekturou plnící různé výpočty a funkce. Pevným a neměnným aspektem výpočetní architektury je její konečný soubor základních operací, které je **hardware** počítače schopen vykonat. Základem flexibility počítače je pak jeho schopnost tyto základní operace různým způsobem kombinovat a řetězit na základě programových instrukcí (**softwaru**) uložených v paměti, a tak realizovat různé výpočty různých výpočetních zařízení majících různou výpočetní architekturu. Důležité je, že digitální počítače dokážou napodobit jakoukoli funkci jakéhokoli jiného výpočetního zařízení. Tzn. že všechny digitální počítače jsou konkrétní fyzickou realizací Univerzálního Turingova stroje, který je schopen dokonale napodobit chování každého Turingova stroje (a realizovat tak každou funkci, kterou takový Turingův stroj dokáže spočítat). Jediné, co k tomu UTS potřebuje, je paměť, ve které by byl zaznamenán odpovídající popis chování jiného Turingova stroje. Tento popis chování jiného Turingova stroje není de facto nic jiného než program – sada instrukcí, které UTS říká, jak se má chovat, resp. jak má použít soubor svých základních operací, aby napodobil chování jiného Turingova stroje, který vykonává určitou specifickou funkci. Stejným způsobem je možné se dívat i na programy (software) digitálních počítačů – tedy buď jako na určitou specifickou posloupnost primitivních instrukcí protékajících registrem instrukcí CPU nebo jako na popis nějakého jiného výpočetního stroje s výpočetní architekturou, která mu umožňuje plnit určitou specifickou funkci. Tyto stroje se přitom neskládají z vodičů či tranzistorů, ale z řady instrukcí, které umožňují hardwaru dočasně se chovat způsobem, který odpovídá chování jiného výpočetního stroje. V počítačové vědě se tato zařízení nazývají **virtuální stroje**. Tyto virtuální stroje mají svojí vlastní virtuální architekturu, pracují s vlastními virtuálními objekty a prostřednictvím svých vlastních virtuálních operací. Takto může například běžný uživatel PC prostřednictvím jistého virtuálního stroje - textového editoru – různě upravovat bloky textu; s pomocí jiného virtuálního stroje – tabulkového procesoru – zase může různým způsobem zpracovávat numerická data; na většině osobních počítačů existuje také jeden speciální virtuální stroj –

operační systém -, který uživateli umožňuje pohodlným a intuitivním způsobem ovládat a spravovat ostatní virtuální stroje nainstalované na počítači. S virtuálním strojem pracuje také programátor, když s pomocí nějakého vysokoúrovňového programovacího jazyka instruuje hardware, jakým způsobem se má chovat, čímž jenom vytváří další virtuální stroj, který je o jednu vrstvu dál od hardwaru než programovací jazyk, ve kterém je virtuální stroj implementován (Dennett, 1993). Tímto způsobem je možné virtuální stroje do sebe téměř neomezeně vkládat a nořit – von Neumannovský stroj (hardware) se nejdříve přemění v nějaký jiný, virtuální stroj, v rámci tohoto virtuálního stroje se implementuje další virtuální stroj, v rámci tohoto pak další atd. Na funkčnost virtuálního stroje přitom nemá vůbec žádný vliv počet vrstev, které ho dělí od hardwarové úrovně. O to, jakým způsobem je daný virtuální



Obrázek 29: Noření virtuálních strojů. von Neumannovský stroj disponuje dostatečně univerzální sadou základních operací, které – jsou-li správně zkombinovány – umožňují implementovat jakýkoli dostatečně dobře specifikovaný výpočet, resp. virtuální stroj, který tento výpočet provádí. Tyto virtuální stroje – mají-li tu správnou funkční architekturu – pak lze použít k implementaci dalších virtuálních strojů. Toto noření se různých virtuálních strojů do sebe lze přirovnat k situaci, kdy se člověk snaží věrně napodobit pohyby člověka, který napodobuje pohyby jiného člověka napodobujícího pohyby dalšího člověka...

stroj implementován na úrovni hardwaru, tedy na úrovni von Neumannovského stroje, se přitom člověk (většinou) nemusí vůbec starat; konkrétní detaily hardwarové úrovně nejsou nijak důležité - alespoň tedy dokud výpočetní profil hardwaru umožňuje a podporuje funkční architekturu virtuálního stroje, který je na něm implementován; a protože von Neumannovský

stroj je ztělesněním Univerzálního Turingova stroje, je takto možné jeho prostřednictvím implementovat jakýkoli jiný výpočetní stroj (pokud mu dodáme - třeba i zprostředkovaně přes několik vrstev dalších virtuálních strojů - odpovídající popis jeho chování, tedy ten správný program). Tato ohromná flexibilita digitálního počítače spolu s předpokladem, že lidská mysl je druhem výpočtu, tak logicky vede k závěru, že s pomocí digitálního počítače (který dokáže provést jakýkoli výpočet) je teoreticky možné implementovat virtuální stroj, jehož funkční architektura by umožňovala realizovat mentální výpočet, tedy výpočet, který je lidskou myslí. Kognitivní vědci takto se vznikem digitálního počítače získali nástroj, s jehož pomocí se mohou pokusit modelovat kognitivní procesy lidské mysli nebo vytvářet (virtuální) stroje s vlastním typem („umělé“) inteligence.

Jak již zde bylo jednou poznamenáno, při tomto modelování mentálních procesů se kognitivní věda zajímá především o abstraktní specifikaci mentálního výpočtu a již mnohem méně pak o konkrétní způsob fyzické implementace takového výpočtu – neboť ten může být realizován mnoha různými způsoby. Ve filozofii mysli se tento postoj – v kognitivní vědě zcela převládající – nazývá **funkcionalismus**: To, co dělá mysl myslí, není to, z čeho je sestavena, tedy její substance, ale to, co dělá, tedy její funkce. A podobně jako není pro funkci stolu důležité, zda je ze dřeva nebo ze železa, není ani pro funkci mysli podstatné to, zda její substanci tvoří nějaký biologický materiál nebo křemíkové čipy. Podstata nějakého mentálního stavu (například určitého přesvědčení, obavy nebo přání) takto nevězí v jeho biologickém substrátu, ale v jeho roli, kterou hraje v celkové „ekonomice“ mentálního života; a tuto roli mohou hrát (v tom správném „ekonomickém systému“) různé fyzikální stavy.

Tento poměrně volný vztah mezi myslí a mozkiem se často přirovnává ke vztahu mezi počítačovým programem a hardwarem: Počítačový program není na určité úrovni abstrakce nic jiného než soubor několika základních operací, které jsou aplikovány na (určitým způsobem symbolicky reprezentované) vstupní informace a které tyto vstupní informace transformují na výstupní informace (symbolicky reprezentované v obdobném formátu); jedná se de facto o specifikaci Turingova stroje, který dokáže realizovat daný typ výpočtu. Takto definovaný program pak lze zapsat v řadě různých programovacích jazyků jako je LISP, BASIC, Java nebo PASCAL a běžet může na řadě různých strojů s odlišným hardwarem. Podle zastánců funkcionalismu je lidská mysl k mozku ve stejném vztahu jako software k hardwaru. Nebo jinak řečeno, **lidská mysl je software, který běží na hardwaru lidského mozku**. Tato nezávislost mentálního softwaru na mozkovém hardwaru však není v žádném případě absolutní, neboť funkce vždy klade na povahu substance, která ji implementuje, určité minimální požadavky, které musí být splněny, aby daná funkce mohla být vůbec realizována.

Takto je například jen velice obtížné si představit, že by pila, která by byla z papíru místo ze železa, mohla dál úspěšně plnit svou obvyklou funkci. Při specifikaci mentálního výpočtu je důležitým omezením především ta skutečnost, že mysl se vždy nachází v těle nějakého biologického organismu, a že její hlavní funkcí tedy není řešení nějakých abstraktních úloh a problémů, ale především řízení a koordinace různých tělesných činností, které musí být přesně načasovány a sladěny s dynamikou vnějšího prostředí. To spolu s poměrně pomalým a neohrabaným biologickým hardwarem¹ nejspíše přimělo evoluční algoritmus matky přírody k selekci a exploataci méně tradičních a méně elegantních výpočetních strategií, které by asi lidskému inženýrovi - navrhujícímu design kognitivní architektury umělého organismu - nepřišly vůbec na mysl, neboť by se při svém návrhu nemusel potýkat se stejnými omezeními jako příroda². Tzn. že při specifikaci mentálního výpočtu se kognitivní věda zřejmě neobejde bez detailních poznatků týkajících se vlastností jeho biologického substrátu, tedy mozku (**výpočetní neurověda** - *computational neuroscience*), a bez poznatků o struktuře a dynamice vnitřního i vnějšího prostředí, v jehož kontextu se daný výpočet odehrává (**vtělená** či **tělesná kognice** - *embodied cognition* a **distribuovaná kognice** - *distributed cognition*).

1.2.3 Výpočetní teorie mysli

Cílem kognitivní vědy je vytvořit materialistickou teorii, která by dokázala propojit doménu, kterou popisuje folková psychologie, s doménou, kterou má na starosti neuroanatomie a neurofyzologie. Jinými slovy, cílem je najít způsob, jak propojit mysl - (obsahující různé mentální entity s vlastní intencionalitou/sémantikou) s neurálními funkcemi a strukturami (které žádnou intencionalitu nemají). Většina kognitivních vědců (s výjimkami jako je *John Searle* nebo *Hubert Dreyfus*) se shodne na tom, že teorie, která to nejspíše dokáže, bude teorií výpočetní – v té či oné podobě. Tzn. že většina kognitivních vědců souhlasí s tím, že lidská mysl je druh výpočetního stroje/software, který je tím či oním

¹ Zatímco rychlost přenosu elektrochemického signálu v nervové soustavě může dosáhnout maximální hodnoty okolo 100 m/s, v křemíkových čípech je rychlost přenosu elektrických signálů mnohonásobně vyšší. Takže zatímco mozkovou činnost „měříme v řádech milisekund (1/1000 sekundy), činnost počítačů... musíme měřit v řádech nanosekund (1/1 000 000 000 sekundy)“ (Cumminsová, 1998, s. 39).

² Jak známo, evoluce přírodním výběrem při „návrhu“ designu živých organismů vždy pracuje pouze s tím, co již existuje a co má k dispozici; trochu si s tím pohraje, ale nikdy to nemůže úplně od základu předělat. Evoluci je možné takto přirovnat k přestavbě lodi během plavby na širém moři, kde lze k úpravám a vylepšením použít pouze ten materiál, který se již na lodi nachází. Navíc se evoluce při svých projektech nemůže spoléhat na žádnou schopnost předvídavosti, která by jí umožňovala anticipovat možné potíže a problémy s navrhovaným designem a přicházet tak s kvalitnějšími návrhy. To jediné, co evoluce dělá, je že upřednostňuje ty náhodné změny, které organismu pomáhají v daném typu prostředí přežít a úspěšně se rozmnožovat. „Navržené“ struktury tak jsou výsledkem nekonečné řady dílčích experimentů majících povahu pokusu a omylu.

způsobem implementován v biologickém hardwaru lidského mozku. Ovšem tato shoda na výpočetní povaze lidské mysli platí pouze na té nejjobecnější rovině, kde je lidská mysl definována jako specifickým způsobem koordinovaný proces mechanické manipulace se syntaktickými stavy, který dokáže zachovat jejich sémantické charakteristiky jakožto systematicky interpretovatelných objektů reprezentujících nějaké jiné entity. Takové tvrzení o výpočetní povaze lidské mysli je však příliš obecné a vejde se pod něj řada různých názorů na konkrétní povahu funkční architektury výpočetního stroje, který mentální výpočet realizuje. Jinými slovy, kognitivní vědci se mohou lišit v názoru na to, jaký je soubor základních operací takové kognitivní výpočetní architektury, a v názoru na to, jaká je povaha entit, na které jsou tyto základní operace aplikovány. Odhlédneme-li od množství různých výpočetních modelů řady dílčích kognitivních funkcí (zraková percepce, paměť, řešení problémů, analogické myšlení, jazyková kompetence...), můžeme všechny výpočetní modely rozdělit do dvou hlavních skupin podle toho, zda při modelování vybrané kognitivní funkce postulují existenci výpočetních operací na symbolové, nebo subsymbolové úrovni. K objasnění rozdílů mezi těmito dvěma typy kognitivních systémů je užitečný Clarkův (1989, 2001) koncept sémanticky transparentní systémů.

1.2.3.1 Sémanticky transparentní systémy

Koncept sémanticky transparentní systémů je asi nejlépe uchopitelný z perspektivy, kterou nabízí Marrovy tři základní úrovně analýzy, na kterých by měl být popsán každý kognitivní systém, kterému chceme plně porozumět. Tyto tři úrovně analýzy jsou již detailně popsány na straně 28; zde proto uvádím pouze stručné shrnutí: Na první (výpočetní) úrovni analýzy je specifikována funkce, kterou daný kognitivní systém počítá, na druhé (reprezentační a algoritmické) úrovni je specifikována posloupnost výpočetních operací a způsob reprezentace vstupních a výstupních informací, na které jsou tyto operace aplikovány, a nakonec na třetí (implementační) úrovni je popsán způsob, jak může být celý výpočet realizován nějakým fyzickým zařízením. Popis funkce, kterou kognitivní systém vykonává (1. úroveň), je obvykle formulován v pojmech, které člověk celkem běžně používá ve svém myšlení a jazyce, když se snaží uchopit nějakou problémovou doménu (proto se této úrovni analýzy také někdy říká pojmová). Tzn. že kdybychom analyzovaly například nějaký program, který by měl za úkol na základě vstupních lékařských dat určit diagnózu pacienta, ve specifikaci jeho funkce by se asi objevila taková běžná slova jako je „krevní tlak“, „hladina cukru“ nebo „tělesná teplota“.

Sémanticky transparentní systém je pak takový systém, jehož algoritmická a

reprezentační úroveň popisu těsně koresponduje s jeho pojmovou úrovní analýzy. Z toho vyplývá, že **výpočetní operace** (specifikované na algoritmické a reprezentační úrovni analýzy) **jsou v sémanticky transparentním systému aplikovány na reprezentační struktury, které jsou izomorfní s pojmovými entitami na pojmové úrovni popisu systému.** Řekněme, že jedna ze specifikovaných funkcí lékařského diagnostického programu bude spočívat ve schopnosti vyvodit z přítomnosti vysoké teploty, kašle a pocení závěr, že člověk má angínu, nebo chřipku. Tuto funkci můžeme na pojmové úrovni specifikovat pomocí jednoduchého transformačního pravidla:

JESTLIŽE (vysoké teplota a kašel a pocení) POTOM (angína nebo chřipka)

Aby bylo možné diagnostický program považovat za sémanticky transparentní, objekty jako „vysoké teplota a kašel a pocení“ či „angína nebo chřipka“, které jsou na pojmové úrovni předmětem transformačního pravidla (*JESTLIŽE-PAK*), musí mít svou jasnou strukturální analogii na algoritmické úrovni; ta tedy musí zahrnovat výpočetní operace, které jsou definovány s ohledem na strukturu symbolických reprezentací, která je více či méně totožná se strukturou pojmových výrazů jako je „vysoké teplota a kašel a pocení“.

Výpočetní architektura sémanticky transparentních systémů takto předpokládá existenci reprezentačního média s dosti specifickou syntaxí, která věrně odráží sémantickou strukturu pojmové úrovně analýzy kognitivního systému. Toto specifické reprezentační médium se nazývá **jazyk myšlení** (*language of thought*) nebo také **mentalíza** (*mentalese*) (Fodor, Pylyshyn, 1988). Jeho základní charakteristikou je kombinatorická povaha, tzn. že se skládá z hierarchie reprezentačních/symbolových atomů, které se systematickým způsobem kombinují v různé komplexní reprezentační/symbolové struktury (molekuly), jejichž sémantická interpretace je systematickou funkcí sémantické interpretace jejich atomických reprezentačních/symbolových komponent. O všech reprezentačních/symbolových prvcích jazyka myšlení (atomických i těch molekulárních) lze přitom říci, že odkazují k běžným pojmům jako například „zítra“, „pes“ „podstatné jméno“ nebo „zelený“. Výpočetní operace sémanticky transparentního systému jsou pak vždy definovány ve vztahu k této specifické struktuře reprezentací jazyka myšlení.

Předpoklad existence takového jazyka myšlení a mentálních operací zohledňujících jeho specifickou syntaktickou a sémantickou strukturu značně usnadňuje konstrukci nebo explanaci systémů, jejichž činnost lze systematicky interpretovat v souladu se sémantikou nějaké problémové domény, neboť jestliže sémantická struktura problémové domény úzce

koresponduje se syntaktickou strukturou symbolické reprezentace, potom lze celkem snadno nahlédnout, jakým způsobem daný systém dokáže prostřednictvím čistě syntaktických vlastností symbolických reprezentací napodobit určité sémantické zákonitosti. Například u šachového programu, jehož výpočetní operace jsou definovány na symbolických strukturách, které přímočaře označují pojmy jako „šach-mat“, „dáma“ nebo „pěšec“, je celkem zřejmé, že jestliže prostřednictvím takto definovaných výpočetních operací donutím symboly chovat se určitým specifickým způsobem, výsledkem bude inteligentní chování hráče šachu, který dokáže rozumně uvažovat o možných herních tazích. Sémanticky transparentní systémy, které takto vycházejí z předpokladu, že výpočetní procesy vysvětlující kognitivní procesy probíhají na úrovni blízké sémantické doméně běžného jazyka psychologie, se nazývají **symbolové systémy**. Teoretickou artikulací názoru, že výpočetní architektura takových symbolových systémů je dostatečná k implementaci obecného inteligentního chování (tedy inteligence lidského typu), je Newellova a Simonova **hypotéza fyzikálního symbolového systému**.

1.2.3.2 Symbolové systémy

Fyzikální symbolový systém je podle Newella a Simona (1976, s. 40-42) jakékoli fyzické zařízení, které obsahuje:

- 1) soubor interpretovatelných a kombinovatelných symbolů (implementovaných jako určité fyzikální stavy),
- 2) sadu základních operací (jako je kopírování, spojování, přesouvání, vytváření nebo odstraňování), které lze na tyto symboly (v souladu s určitými instrukcemi) aplikovat,
- 3) a které je umístěno v širší síti reálných objektů.¹

Fyzikální symbolový systém je tedy jakýkoli systém, ve kterém lze určitým (specifickým způsobem manipulovatelným) symbolům připsat arbitrární významy, a který lze prostřednictvím pečlivého naprogramování přimět chovat se způsobem konzistentním s takto arbitrárně připsanými významy. Důležité je, že podle Newella a Simona (1976, s. 86) „fyzikální symbolový systém má všechny nutné a dostačující prostředky pro implementaci

¹ A physical symbol system „consists of a set of entities, called symbols, which are physical patterns that can occur as components of another type of entity called an expression (or symbol structure). Thus, a symbol structure is composed of a number of instances (or tokens) of symbols related in some physical way (such as one token being next to another). At any instant of time the system will contain a collection of these symbol structures. Besides these structures, the system also contains a collection of processes that operate on expressions to produce other expressions: processes of creation, modification, reproduction and destruction. A physical symbol system is a machine that produces through time an evolving collection of symbol structures. Such a system exists in a world of objects wider than just these symbolic expressions themselves” (Newell, Simon, 1976).

*obecné intelligence*¹ (tedy ten typ intelligence, který lze pozorovat u člověka). To tedy potom znamená, že **každý fyzikální systém, který je schopen vykazovat obecnou inteligenci, musí být fyzikálním symbolovým systémem, a že každý fyzikální symbolový systém může být naprogramován tak, aby vykazoval obecnou inteligenci.** Toto tvrzení je přitom možné interpretovat dvěma různými způsoby (Clark 1989): Ten první a nesprávný je ten, 1) že fyzikální symbolový systém je univerzální stroj, který dokáže realizovat jakýkoli dostatečně dobře specifikovaný výpočet, 2) že mysl je ve své podstatě druh výpočtu a 3) že inteligenci lze realizovat prostřednictvím univerzálního stroje (tj. také prostřednictvím fyzikálního symbolového systému). Takto chápáné je však Newellovo a Simonovo tvrzení příliš obecné. Newell se Simonem ve skutečnosti svou hypotézu chápou jako specifické a empiricky testovatelné (resp. popperovsky vyvratitelné) tvrzení o konkrétní povaze výpočetní architektury virtuálního stroje (fyzikálního symbolového systému), který je podkladem obecné intelligence. Takto úžeji chápáná hypotéza fyzikálního symbolového systému potom tedy tvrdí, že **virtuální stroj provádějící von Neumannovské operace na standardních symbolických atomech má přímé, nutné a dostačující prostředky k implementaci obecné intelligence:**

Odkaz na standardní symbolické atomy zde poukazuje na povahu entit/symbolických reprezentací, na které jsou aplikovány výpočetní operace fyzikálního symbolového systému. V tomto případě to jsou symbolické reprezentace, jejichž jednotlivé prvky (symbolové atomy) lze přímočaře a intuitivně interpretovat v souladu s pojmy běžného diskurzu. Tzn. že fyzikální symbolový systém představuje sémanticky transparentní systém (viz s. 69-71). Tímto Newell se Simonem jednoznačně identifikují úroveň abstrakce, která je podle nich vhodná k popisu lidské mysli jako výpočtu.

Zmínka o von Neumannovských operacích odkazuje na výpočetní (symboly manipulující a transformující) operace, které jsou sériově (postupně) aplikovány na standardní symbolové atomy. Mezi tyto operace, které má fyzikální symbolový systém k dispozici, patří kopírování symbolů, spojování symbolů, přepisování symbolů, čtení symbolů, jejich identifikace či porovnávání (*pattern-matching*). Všechny tyto základní operace lze nalézt také u běžného von Neumannovského stroje, což je v souladu s původem jeho výpočetní architektury, která je přes Turingův stroj odvozena ze základních operací lidské mysli, které Alan Turing pečlivou introspekci vyextrahoval z proudu svého vědomí během řešení matematických úloh. Stejně operace lze nalézt také v LISPu - oblíbeném programovacím jazyku kognitivních vědců, kteří

¹ „A physical symbol system has the necessary and sufficient means for general intelligent action.“

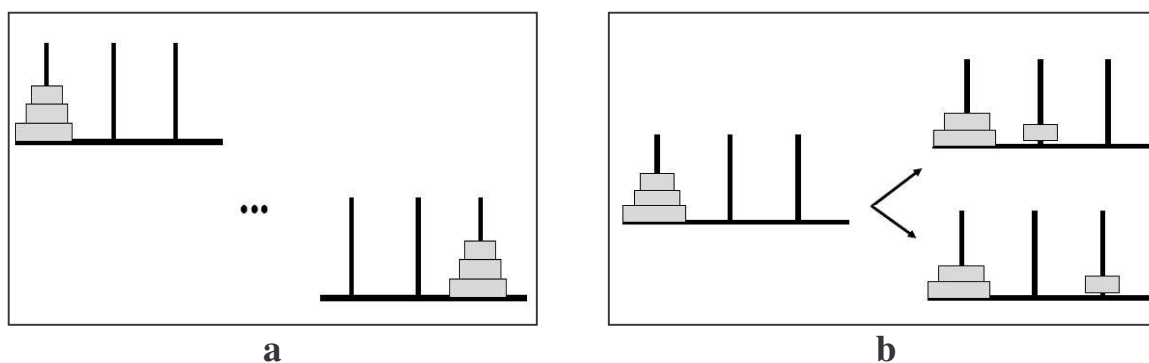
jej často používají k modelování některých (zejména vyšších, centrálních) kognitivních procesů jako je např. řešení problémů (*problem solving*).

Fyzikální symbolový systém je virtuální stroj, který je implementován na biologickém hardwaru lidského mozku, jehož výpočetní architektura (tzn. soubor základních operací a reprezentačních struktur, na které jsou tyto operace aplikovány) se zásadním způsobem liší od té, kterou disponuje fyzikální symbolový systém. Neuronální výpočetní stroj však poskytuje dostatečně univerzální sadu operací, která mu umožňuje implementovat výpočetní architekturu fyzikálního symbolového systému. Podobně může být fyzikální symbolový systém implementován i prostřednictvím nějaké jiné výpočetní architektury s odlišnou sadou základních operací; avšak u každého systému, který je schopen inteligentního jednání, musí – podle Newella a Simona - na určité úrovni popisu existovat symbolový systém, který je nutnou a dostatečnou podmínkou jakéhokoli inteligentního jednání. Z toho všeho vyplývá, že o výpočetní profil biologického hardwaru se nemusíme vůbec starat; důležitá je pouze vyšší úroveň, tj. virtuální stroj fyzikálního symbolového systému s jeho vlastní sadou základních operací a objektů, na které jsou tyto operace aplikovány. Je to obdoba situace, kdy se programátor nebo uživatel nezajímá o strojový kód svého von Neumannovského stroje (PC), který mu leží na stole, ale pouze o operace a objekty, které má k dispozici na vyšší úrovni virtuálního stroje nějakého programovacího jazyka nebo uživatelského programu.

Poslední zmínka o přímých prostředcích k implementaci obecné inteligence pak znamená, že fyzikální symbolový systém je schopen toto chování generovat bezprostředně – tzn. bez ohledu na to, na jakém skutečném nebo virtuálním stroji je simulován, a aniž by musel simulovat jiný virtuální stroj s výpočetní архитектурou odlišnou od té jeho vlastní. Výpočetní architektura fyzikálního symbolového systému (soubor jeho základních, von Neumannovských operací postupně aplikovaných na standardní symbolové atomy) takto podle Newella a Simona podporuje všechny možné aspekty lidské (obecné) inteligence. Newellova a Simona teorie fyzikálního symbolového systému je takto formulací nové „mentální chemie“, která popisuje lidskou mysl jako druh počítačového programu: „*Atomy této mentální chemie jsou symboly, které jsou kombinovatelné do větších a složitějších asociačních struktur... Základní „reakce“ mentální chemie využívají elementárních informačních procesů, které operují na jednotlivých symbolech a symbolických strukturách: kopírování symbolů, vyvolávání symbolů, vkládání a vytváření symbolů a porovnávání symbolů.*“¹ (Simon, 1979, s. 363)

¹ „*The atoms of this mental chemistry are symbols, which are combinable into larger and more complex associational structures... The fundamental „reactions“ of the mental chemistry employ elementary information*

Konkrétním příkladem aplikace výpočetní architektury symbolového systému k implementaci určitého aspektu inteligentního chování člověka je **produkční systém** použitý Newellem a Simonem (1972) k modelování kognitivních procesů spojených se schopností řešit problémy (*problem solving ability*). Na základě analýzy záznamů hlasitého přemýšlení lidí řešících různé, většinou dobře strukturované problémy, přišli Newell se Simonem s ideou, že řešení problémů obecně spočívá v **heuristickém prohledávání symbolického problémového prostoru** (*symbolic problem space*): Problémový prostor je druh **stavového prostoru problémové situace**, kterou je možné popsat prostřednictvím řady různých parametrů, které definují její **aktuální stav**. **Množina všech možných aktuálních stavů** problémové situace pak tvoří stavový prostor daného problému. Hodnoty jednotlivých parametrů problémové situace lze přitom chápat jako souřadnice, které lokalizují aktuální stav problému v n -rozměrném prostoru, kde n se rovná počtu parametrů charakterizujících danou problémovou situaci. Koncept stavového prostoru problému lze dobře ilustrovat na relativně jednoduché úloze, kterou představuje *Hanojská věž* (*Tower of Hanoi*) – původně hlavolam vytvořený na konci 19. století francouzským matematikem *Eduardem Lucasem*, později nástroj výzkumu různých kognitivních funkcí (paměť, učení, řešení problémů) a v současnosti jeden z neuropsychologických nástrojů používaných k měření tzv. exekutivních (dys)funkcí, zejména schopnosti plánovat, realizovat, monitorovat a adaptovat posloupnosti činností za účelem dosažení předem stanoveného cíle (Brenkus, 2005). Při práci s klasickou verzí Hanojské věže¹ je úkolem testované osoby přemístit s pomocí co nejméně tahů a v co

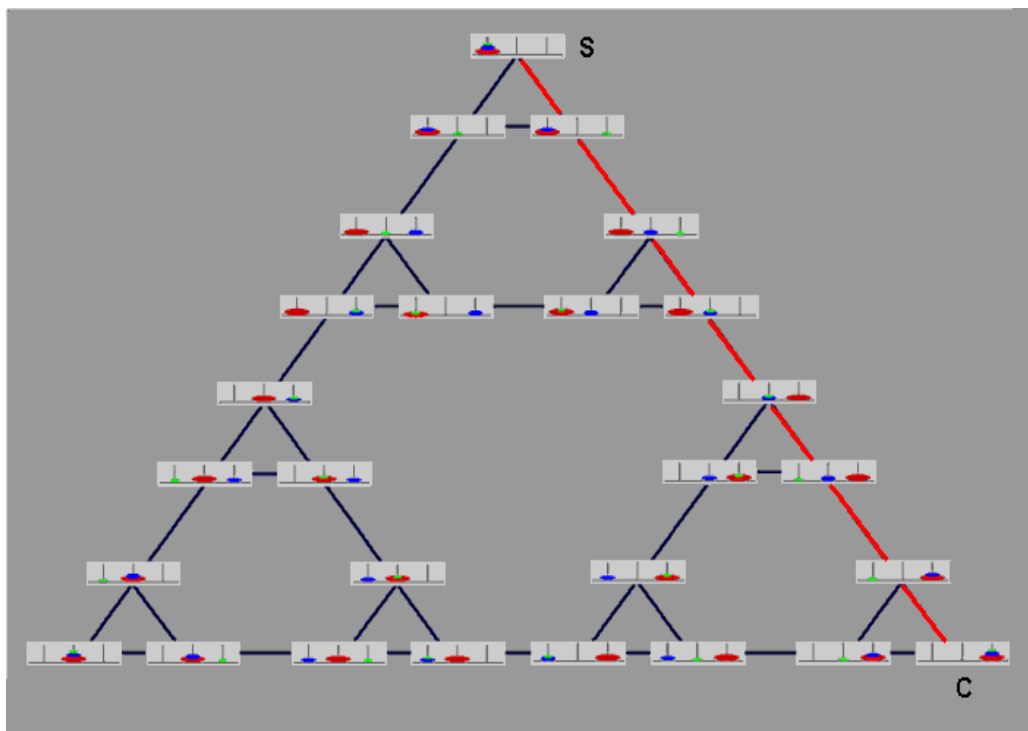


Obrázek 30: Hanojská věž. a) V klasické verzi Hanojské věže je cílem přemístit v co možná nejmenším počtu tahů věž ze tří disků na levém kolíku na pravý kolík. b) Při tomto přesouvání se člověk musí držet určitých pravidel, která definují povolené tahy. (převzato z <http://hci.ucsd.edu/102a/05-lectures/W3a.pdf>)

processes that operate upon symbols and symbol structures: copying symbols, storing symbols, retrieving symbols, inputting and outputting symbols, and comparing symbols.“

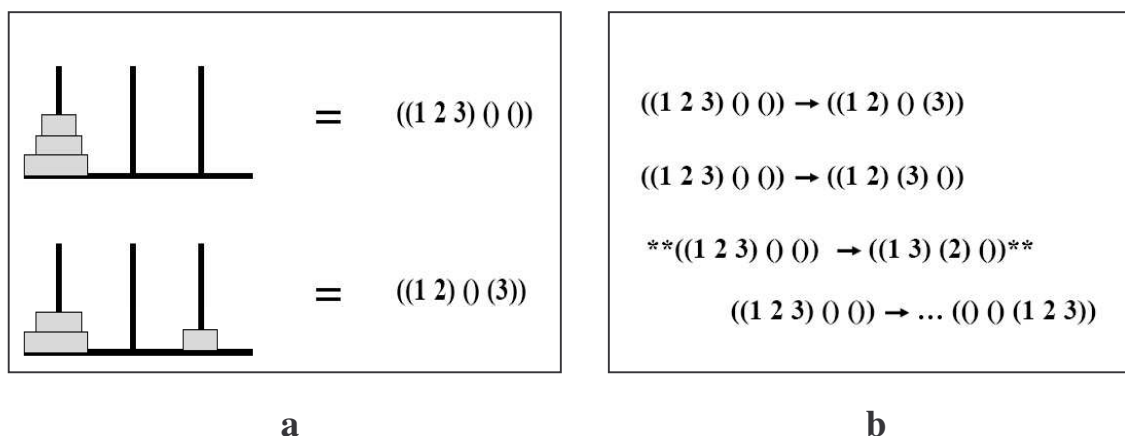
¹ Hanojská věž je pouze jedním z představitelů širší skupiny tzv. *věžových testů*, které se mohou lišit počtem kotoučů, způsobem fyzické implementace (například druhem přesouvaných objektů, jejich barevným či jiným rozlišením...) a pravidly, která definují povolené tahy.

nejkratším časovém limitu věž ze tří kotoučů z výchozí pozice na levém kolíku do cílové pozice na pravém kolíku za použití prostředního kolíku, který slouží jako pracovní prostor; v jednom tahu přitom nesmí proband vzít do ruky více než jeden kotouč a nikdy nesmí položit větší kotouč na menší. Stavový prostor testu Hanojské věže sestává ze všech možných konfigurací tří kotoučů na třech kolících. Vzhledem k malému množství možných stavů problémové situace lze celý stavový prostor zachytit prostřednictvím nevelkého herního stromu (viz obrázek 31) obsahujícího 27 různých stavů, z nichž každý odpovídá jedné z možných konfigurací kotoučů; spoje mezi jednotlivými stavy představují pravidly povolené transformace jednoho stavu na druhý. Úkolem testované osoby je pak najít v tomto stavovém prostoru nejkratší možnou trajektorii (na obrázku 31 vyznačenou červenou linkou) vedoucí od počátečního stavu (S) přes několik mezilehlých stavů až k cílovému stavu (C), který představuje řešení celého problému. Podle Newella a Simona se člověk snaží tuto trajektorii



Obrázek 31: *Problémový prostor testu Hanojská věž.*

najít tím způsobem, že prohledává symbolický problémový prostor, tj. symbolicky kódovaný stavový prostor problémové situace. Takové prohledávání symbolického problémového prostoru přitom probíhá tak, že člověk nejdříve vytvoří symbolický stav, který reprezentuje výchozí a cílovou situaci (tj. problém), a potom se snaží prostřednictvím aplikace různých operátorů (transformačních operací, které lze na různé stavy aplikovat a transformovat je tak



Obrázek 32: a) Symbolická reprezentace jednotlivých stavů problémové situace a b) povolených a zakázaných operací aplikovaných na tyto stavy (spolu s cílovým stavem). (převzato z <http://hci.ucsd.edu/102a/05-lectures/W3a.pdf>)

na další stavy) transformovat počáteční stav na stav cílový.¹ Vzhledem k tomu, že stavový prostor většiny problémů je kvůli **kombinatorické explozi** možností (která vzrůstá exponenciálně s počtem prvků, jejichž kombinace definuje stavový prostor problému²) příliš rozsáhlý na to, aby jej bylo možné v rozumném čase celý systematicky prohledat, používají lidé při jeho prohledávání většinou nějakou **heuristiku**³, která jim sice nezaručuje, že se doberou řešení problému a že se ho doberou tím nejefektivnějším možným způsobem, nicméně člověk, který se takovou heuristikou řídí, se dokáže poměrně snadno a rychle dostat do té oblasti problémového prostoru, kde může s trochou štěstí na řešení problému narazit. Newell se Simonem - na základě analýzy záznamů hlasitého přemýšlení lidí řešících různé problémy - několik takových heuristik popsali; takto identifikovali například *analýzu prostředků a cílů* (*means-ends analysis* – strategie spočívající v průběžném monitorování rozdílu mezi aktuálním a cílovým stavem a v realizaci činností, které tento rozdíl zmenšují),

¹ Je to jakási obdoba jednoduchého slovního hlavolamu (tzv. „slovního řetězu“) spočívajícího ve změně jednoho slova na druhé prostřednictvím postupné obměny jednotlivých hlásek, s tím omezením, že všechna slova, která vzniknou v každém jednotlivém kroku obměny jedné hlásky, musí být smysluplným podstatným jménem v 1. pádě jednotného čísla (například **rys** – ryk – rok – lok – lom – lem – ley – led – **med**).

² Klasickým příkladem úlohy „trpící“ kombinatorickou explozí jsou šachy nebo problém obchodního cestujícího.

³ **Heuristika** obecně představuje krátkou a relativně snadnou proceduru, která obvykle směřuje ke správné odpovědi, ale neručí za to, na rozdíl od **algoritmu**, který představuje nikdy neselhávající postup, který zaručeně dosáhne stanoveného cíle, ovšem často za cenu vysoké složitosti a časové náročnosti provedení daného postupu, což je důvod, proč lidé při řešení svých každodenních problémů sahají spíše po *suboptimálních heuristikách*, které sice člověku často neumožňují dosáhnout optimálních řešení, ale většinou mu umožňují dosáhnout řešení, která jsou uspokojivě z hlediska jím zvolených kritérií a z hlediska poměru „cena“ (složitost a časová náročnost procedury)–„výkon“ (přesnost procedury). Tento přístup k řešení každodenních problémů vyplývá z toho, že člověk obvykle nemá k dispozici neomezené množství času, informací a výpočetní kapacity, aby při usuzování a rozhodování zvažoval všechny možné alternativy ze všech možných relevantních hledisek - viz **teorie omezené racionality** Herberta Simona (Simon, 1982; Kahneman, Slovic, Tversky, 1982; Gigerenzer, Todd a kol., 1999; Gilovich, Griffin, Kahneman, 2002).

vyhledávání extrému (hill-climbing – postup spočívající ve výběru těch stavů, které se nejvíce podobají stavu cílovému), *generování a testování nápadů (generate-and-test* – strategie spočívající v produkování různých postupů a následném ověřování jejich účinnosti), nebo *dopředný/zpětný postup (forward/backtracking* - strategie spočívající v postupu od začátku směrem ke konci, nebo obráceně).

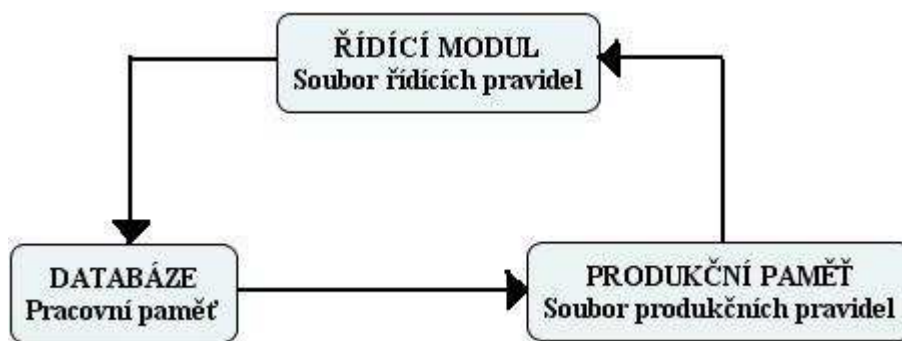
Toto prohledávání problémového prostoru se Newell se Simonem pokusili implementovat a modelovat prostřednictvím symbolové výpočetní architektury **produkčního systému**. Produkční systém se skládá ze tří základních složek: z *databáze*, ze souboru *produkčních pravidel* a z *řídícího modulu*. Databáze produkčního systému funguje jako druh pracovní či krátkodobé paměti, ve které je reprezentován aktuální stav prostředí produkčního systému. V případě řešení testu Hanojské věže by tímto stavem byla aktuální konfigurace kotoučů na kolících hrací desky. Produkční pravidla mají podobu pravidel typu *JESTLIŽE-PAK*, která ve své podmínkové části specifikují různé situace, které když nastanou mají za následek určitou činnost a aktivitu, která je specifikována v důsledkové části produkčního pravidla: *KDYKOLI JE SPLNĚNA určitá podmínka, POTOM PROVEĎ určitou činnost*. V systému řešícím test Hanojské věže by takovým produkčním pravidlem mohlo být například pravidlo „*Jestliže jsou všechny tři kolíky obsazeny, potom přesuň nejmenší kotouč na kolík, kde se nachází kotouč střední velikosti*“. Na tato produkční pravidla se lze dívat buď jako na dlouhodobou deklarativní paměť, ve které jsou uloženy všechny vědomosti produkčního systému o okolním světě, nebo jako na dlouhodobou procedurální (produkční) paměť, ve které je uložen seznam všech jeho možností, jak v tomto světě jednat. Poslední součástí produkčního systému je řídící modul, který na základě heuristik zakódovaných do řídících pravidel určuje, které z produkčních pravidel se použije. Tyto tři základní složky produkčního systému koordinují svou činnost v rámci poměrně jednoduchého algoritmu, který stojí v pozadí fungování produkčního systému jako celku. Tento algoritmus má zhruba následující podobu:

1. **načti** aktuální stav prostředí do pracovní paměti (PP)
2. **dokud** obsah PP nevyhovuje podmínce ukončení činnosti, **prováděj**
3. **začni**
4. **vyber** některé produkční pravidlo *p*, které lze aplikovat na obsah PP
5. **aktualizuj** obsah PP
6. **ukonči**

Produkční systém pracující podle tohoto algoritmu nejdříve načte aktuální stav prostředí do své pracovní paměti; poté co tak učiní, přejde k realizaci procedury, kterou vykonává do té doby, dokud obsah pracovní paměti nesplňuje podmínku, která je nutná k ukončení činnosti

produkčního systému (= dosažení cílového stavu). V rámci této procedury se aktivují všechna produkční pravidla, jejichž podmínková část je naplněna fakty obsaženými v pracovní paměti. Z těchto produkčních pravidel potom řídicí modul (na základě heuristických strategií kódovaných v podobě dalších produkčních pravidel) vybere jedno produkční pravidlo, které se aplikuje na obsah pracovní paměti. Obsah pracovní paměti se tímto změní a zaktualizuje. Celý proces pak začíná nanovo a pokračuje do té doby, dokud není v pracovní paměti zakódováno řešení daného problému.

Jakkoli se může zdát výpočetní architektura produkčního systému triviální a jednoduchá, produkční systém dokáže vypočítat totéž, co dokáže vypočítat jakýkoli Turingův stroj, tzn. vše, co lze vůbec vypočítat. To, že tomu tak je, lze dokázat jednoduchou úvahou týkající se možnosti emulace funkce Univerzálního Turingova stroje (UTS) prostřednictvím produkčního systému. To, že je něco takového možné, naznačuje již skutečnost, že obě dvě výpočetní



Obrázek 33: Struktura výpočetní architektury produkčního systému.

zařízení sdílejí de facto totožnou abstraktní strukturu – u obou můžeme nalézt nějakou paměť a nějaký procesor: UTS – jak již víme - sestává z paměťové pásky, konečného automatu a ze čtecí/zapisovací hlavy: Na pásce UTS je v podobě posloupnosti fixní sady symbolů zakódováno chování nějakého jiného Turingova stroje, který vykonává určitý specifický výpočet. Konečný automat pak prostřednictvím čtecí/zapisovací hlavy může na paměťové pásce vykonávat určité operace: načítat, mazat, přepisovat a zapisovat symboly a pohybovat čtecí/zapisovací hlavou vždy maximálně o jedno políčko pásky doprava nebo doleva. Kromě toho konečný automat také může měnit svůj stav. To vše se děje na základě nějaké základní sady instrukcí, které specifikují vstupně-výstupní funkci, která vstupu v podobě aktuálně načteného symbolu a aktuálního stavu konečného automatu přiděluje výstup v podobě zapsaného nebo vymazaného symbolu, změny stavu konečného automatu a pohybu čtecí/zapisovací hlavy. Produkční systém, který by měl za úkol emulovat funkci takového

UTS, dokáže přesně napodobit jeho chování jednoduše tak, že do své pracovní paměti bude ukládat informace o aktuálně načteném symbolu a o aktuálním stavu konečného automatu, a že ve své produkční paměti bude mít k dispozici sadu produkčních pravidel, která mu umožní měnit obsah své pracovní paměti v souladu se sadou základních instrukcí UTS. Takové produkční pravidlo by mohlo vypadat například následujícím způsobem: *JESTLIŽE stav konečného automatu je A a načteným symbolem je 1, POTOM změn stav konečného automatu na B, vymaž z pásky symbol 1 a posuň čtecí/zapisovací hlavu o 1 políčko pásky doleva.* A protože konečný automat se vždy nachází právě v jednom stavu a čtecí/zapisovací hlava vždy načítá právě jeden symbol, může být splněna podmínková část pouze jednoho produkčního pravidla. Tzn. že řídicí modul může vybírat a volit produkční pravidla na základě velice jednoduchého pravidla: Vyber právě to jediné pravidlo, jehož podmínková část je splněna fakty uloženými v pracovní paměti. Tímto způsobem produkční systém dokáže věrně napodobit funkci UTS. A jestliže UTS dokáže v principu realizovat jakýkoli výpočet, potom to znamená, že jakýkoli výpočet v principu dokáže realizovat i produkční systém – má-li k dispozici tu správnou sadu produkčních pravidel.¹ (Franklin, 1995)

Produkční systémy tvoří teoretický základ řady praktických aplikací v podobě tzv. **pravidlových expertních systémů**, což jsou počítačové systémy, které využívají (v podobě pravidel reprezentované) poznatky formulované experty pro danou problémovou oblast k řešení komplexních úloh a problémů, které jinak vyžadují účast lidského experta. Základními složkami takového expertního systému jsou:

- *báze faktů*, kde jsou uložena fakta charakterizující konkrétní specifický problém, např. že venku právě prší, že daný člověk má silné bolesti hlavy nebo že banka X má 5 % úrokové míry;
- *báze poznatků*, která obsahuje poznatky (v podobě pravidel), které jsou relevantní pro řešení různých úloh z dané problémové domény, např. *Jestliže má dnes podle televizní předpovědi počasí pršet, potom si vezmu deštník*;
- *inferenční mechanismus*, který aplikuje různá meta-pravidla na obsahy báze faktů a poznatků. (Kelemen a kol., 1999)

Pravidlové expertní systémy nacházejí své uplatnění v řadě oblastí jako je např. sestavování počítačových systémů, geologická prospekce, lékařská diagnostika či hodnocení finančních rizik spojených s poskytováním bankovních úvěrů. Všechny tyto expertní systémy jsou

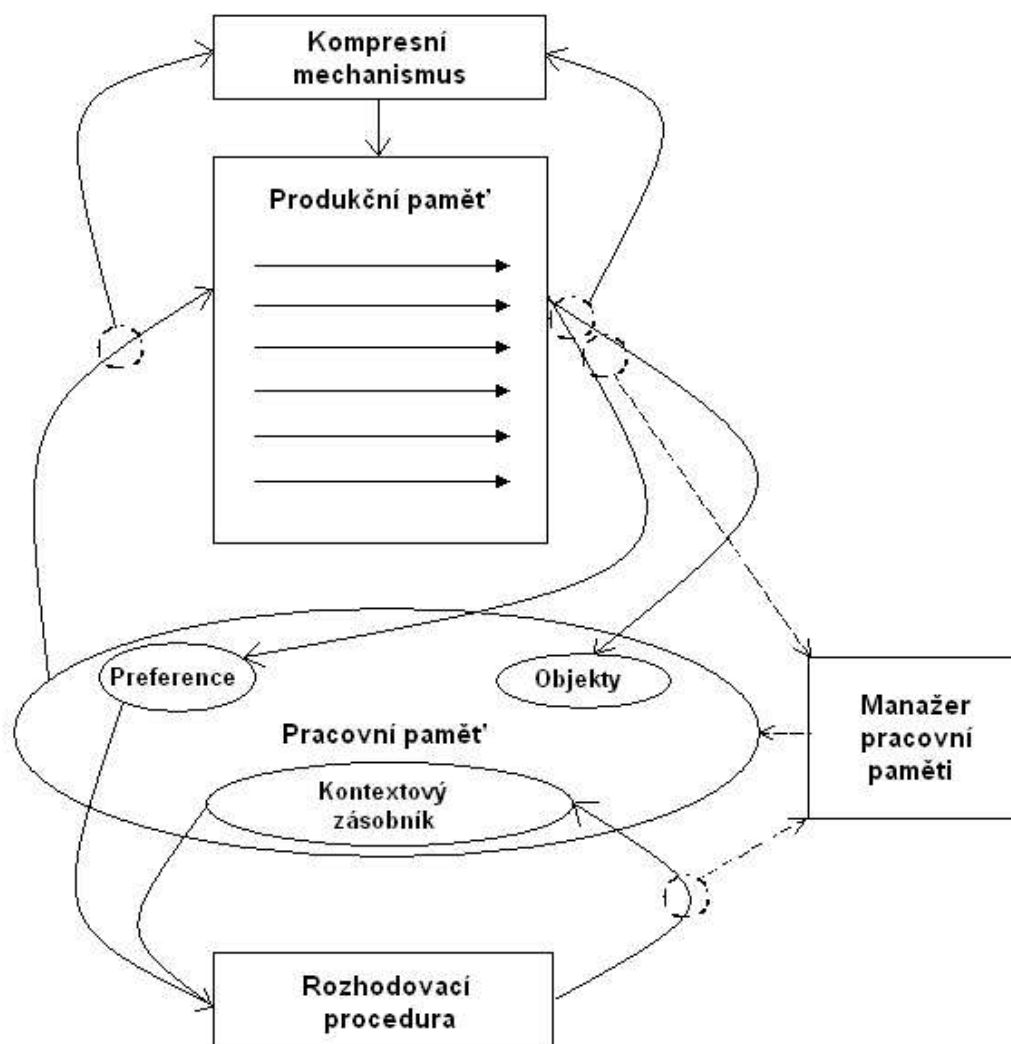
¹ Autorem myšlenky produkčního systému je Emil Post, který přišel s návrhem výpočetní architektury produkčního systému v rámci své snahy formalizovat pojem výpočtu - tedy ze stejného důvodu, ze kterého Alan Turing navrhl své imaginární výpočetní zařízení (Turingův stroj). Není proto žádná náhoda, že Turingův stroj lze simulovat prostřednictvím produkčního systému (a vice versa).

kompetentní pouze ve velmi úzce vymezené problémové oblasti. Podstatu tohoto „fachidiotismu“ dobře vystihuje to, co v jednom televizním interview řekl na adresu počítačů Marvin Minsky, jeden ze spoluzakladatelů oboru umělé inteligence, totiž že počítač *Deep Blue* dokáže vyhrát kdejakou šachovou partii, ale už neví, že se má někam schovat před deštěm, nechce-li zmoknout.

V tomto ohledu představuje světlou výjimku projekt **SOAR** (Rosenbloom a kol., 1992), který v mnohém navazuje na dřívější Newellovu a Simonovu práci v oblasti řešení problémů. Cílem projektu SOAR je vytvořit systém, který by byl plně kompetentní v celé řadě různých oblastí, a který by tak byl schopen vykazovat to, co Newell se Simonem nazývají obecnou inteligencí. SOAR řeší problémy stejně jako běžný produkční systém - tj. prohledáváním symbolického problémového prostoru. SOAR se však liší od běžného produkčního systému – mimo jiné – v tom, že je schopen prohledávat více různých problémových prostorů. SOAR takto při řešení problému vždy nejdříve vybere vhodný problémový prostor, ve kterém bude hledat řešení (cílový stav), v takto vybraném problémovém prostoru potom identifikuje stav odpovídající aktuální a cílové situaci a nakonec vybere a aplikuje na tento stav operátor, který ho změní na další stav atd. až do chvíle, kdy se mu podaří objevit požadované řešení. Veškerá tato činnost (výběr problémového prostoru, výběr operátorů a generování nových stavů v rámci problémového prostoru) se děje na základě znalostí, které má SOAR uloženy v podobě produkčních pravidel ve své dlouhodobé (produkční) paměti.

Produkční pravidla takto hrají v činnosti SOARu stejně klíčovou roli jako v tradičním produkčním systému. To, v čem se SOAR významně liší, je způsob, jak tato produkční pravidla při řešení problémů využívá: V každém okamžiku se SOAR nachází v určitém specifickém *kontextu*, který je definován jeho aktuálním cílem, vybraným problémovým prostorem, aktuálním stavem či pozicí v tomto problémovém prostoru a vybraným operátorem, který se na tento stav použije. Kontext, ve kterém se SOAR právě nachází, je vždy explicitně zakódován v *kontextovém zásobníku* (*context stack*) v pracovní paměti v podobě určitých symbolických objektů (viz obrázek 34, na kterém je zachycena výpočetní architektura SOARu). Produkční pravidla se aktivují na základě obsahu pracovní paměti: nacházejí-li se v ní (symbolické) objekty, které jsou specifikovány v podmínkové části produkčního pravidla, toto pravidlo se aktivuje. Výsledkem aktivace produkčního pravidla je realizace jeho důsledkové části, která vždy specifikuje, jaké (symbolické) objekty se mají přidat do pracovní paměti, nikoli však do kontextového zásobníku, který kóduje aktuální celkový stav SOARu. Na rozdíl od běžného produkčního systému, ve kterém produkční pravidla přímo realizují změnu stavu produkčního systému, v SOARu produkční pravidla

slouží pouze jako jakási asociační paměť, která do pracovní paměti - na základě symbolických objektů v pracovní paměti již obsažených - vyvolává všechny informace (v podobě dalších symbolických objektů), které by mohly být z hlediska daného problému relevantní. Změna celkového stavu SOARu (tj. obsahu jeho kontextového zásobníku v pracovní paměti) tedy probíhá ve dvou rozhodovacích fázích: V první fázi tzv. rozpracování (*elaboration phase*) se všechna produkční pravidla, jejichž podmínková část je naplněna některým z objektů přítomných v pracovní paměti, aktivují a na základě obsahu svých důsledkových částí odesílají do pracovní paměti další objekty, které mohou vyvolat aktivaci dalších produkčních pravidel. Tato aktivita probíhá do té chvíle, dokud do pracovní paměti proudí další nové objekty („vzpomínky“). Na konci této první fáze tak pracovní paměť obsahuje řadu symbolických objektů, které mohou reprezentovat různé problémové prostory, různé stavy v těchto prostorech, různé cílové stavy či různé operátory, z nichž se každý pomyslně hlásí o možnost změnit aktuální stav systému (tj. obsah kontextuálního zásobníku). Vedle těchto objektů se v pracovní paměti na konci fáze rozpracovávání nacházejí také objekty, které se nazývají *preference*. Ty reprezentují informace o žádoucnosti jednotlivých objektů v pracovní paměti a hrají důležitou roli v druhé fázi rozhodovacího cyklu (tzv. *rozhodovací procedura*). Taková preference by mohla například kódovat sdělení, že operátor *x*, který by změnil aktuální stav v problémové prostoru *y* na stav *z*, je akceptovatelný (nebo lepší, horší, nejlepší, nejhorší, nepřijatelný apod.). Rozhodovací procedura pracuje pouze s těmito preferencemi a na jejich základě přesouvá do kontextového zásobníku ty objekty, které z tohoto hodnocení vycházejí jako nejpříjemnější. Po dokončení této rozhodovací procedury se celý rozhodovací cyklus opakuje a to až do chvíle, kdy je nalezeno řešení. Může se stát, že se systém dostane v rámci své rozhodovací procedury do slepé uličky, například proto, že v pracovní paměti nebudou obsaženy žádné preference, které by mohly rozlišit mezi lepší a horší variantou. V takovém případě se automaticky vytváří dílčí cíl spočívající v rozhodnutí o tom, která z možných alternativ by byla nejlepší. Tento nový dílčí cíl se načte do pomyslné svrchní vrstvy kontextového zásobníku, ve kterém je v té chvíli specifikován pouze požadovaný cíl; problémový prostor, ve kterém se bude tento nový cíl hledat, počáteční stav v tomto prostoru a operátor, který by tento výchozí stav nějak transformoval, musí SOAR prostřednictvím svých znalostí (zakódovaných v produkčních pravidlech) teprve vybrat. Kontextový zásobník, ve kterém je vždy zapsána právě aktuální „agenda“ systému, funguje na principu „poslední dovnitř, první ven“; tzn. že po dosažení každého dílčího cíle se aktualizuje obsah kontextového zásobníku do podoby odpovídající cíli, který stojí v hierarchii dílčích cílů o stupeň výš. SOAR takto dokáže promptně a flexibilně vytvářet komplexní hierarchie dílčích



Obrázek 34: Funkční architektura SOARu. (převzato z Laird a kol., 1987)

cílů vedoucích až k cíli hlavnímu. SOAR - stejně jako běžný produkční systém - při prohledávání problémového prostoru využívá různé prohledávací heuristiky, které jsou nepřímou zakódované v produkčních (řídících) pravidlech, v jejichž důsledkové části jsou obsaženy preference. SOAR takto například využívá *analýzu prostředků a cílů*, kdy (na základě přidělených preferencí) vybírá právě ty operátory, které nejvíce redukuje rozdíl mezi aktuálním a cílovým stavem. Poslední důležitou vlastností SOARu je jeho schopnost se učit, která spočívá v jeho schopnosti zkomprimovat objevenou posloupnost kroků (tahů) vedoucích k nějakému dílčímu nebo i konečnému cíli do jediného produkčního pravidla. Takto například SOAR může během rozhodovací procedury narazit na slepou uličku, což může vyvolat potřebu dosáhnout nějakého dílčího cíle; když se SOARu nakonec podaří prostřednictvím prohledávacích heuristik tohoto dílčího cíle dosáhnout, je SOAR schopen – prostřednictvím komprimačního mechanismu (*chunking mechanism*) – vytvořit produkční pravidlo, jehož podmínková část obsahuje stav reprezentující situaci vedoucí k původní slepé uličce a jehož

důsledková část specifikuje v jediném kroku celou sérii tahů opisujících trajektorii vedoucí ke stavu, který vyhovuje danému dílčímu cíli. Tzn. že až SOAR příště narazí na stejnou situaci, která v minulosti vedla do slepé uličky a k potřebě vyřešit nějaký dílčí problém, nebude muset už nic takového řešit, neboť během první fáze rozhodovacího cyklu se bude aktivovat i nové produkční pravidlo a činnost rozhodovací procedury tak bude moci probíhat bez přerušení.

Na základě předcházejícího popisu principů, které stojí v pozadí funkce symbolových systémů, je nyní zřejmé, že tradiční výpočetní modely mentálních procesů založené na těchto principech se vyznačují čtyřmi následujícími charakteristickými vlastnostmi (Clark, 2001): 1) Tyto výpočetní modely používají různé symbolické struktury k explicitnímu kódování všech znalostí, které potřebují k řešení problémů. 2) Na tyto symbolické struktury sériově aplikují základní operace von Neumannovského typu jako je rozpoznávání, kopírování či přesouvání jednotlivých symbolických objektů. 3) Inteligenci modelují jako sériové prohledávání symbolického problémového prostoru, tj. jako „*vytváření a postupné modifikování symbolických struktur dokud se nedosáhne cílové struktury*“¹ (Newell, Simon, 1976, s. 96). 4) Znalosti potřebné k prohledávání problémového prostoru kódují relativně přímočaře z té úrovně myšlení, která je dostupná vědomé introspekci. Tzn. že vycházejí z předpokladu, že podstata lidské inteligence se nachází blízko úrovně relativně pomalého, sériově činného a vědomého myšlení. Hofstadter (1985, s. 632) takto cituje Simona z jedné jeho přednášky, kde údajně pronesl, že „*v kognici se všechno zajímavé děje až za hranicí 100 milisekund – což je čas potřebný k rozpoznání vaší matky*“². Tento předpoklad také zásadním způsobem ovlivňuje výběr modelových úloh a problémů, na kterých se prostřednictvím symbolových systémů modeluje schopnost inteligentní jednání: Jsou to úlohy, které lidé z velké části řeší prostřednictvím pomalého, sekvenčního a vědomého úsudku; tyto úlohy jsou většinou dobře strukturované, tzn. že mají konečný počet dobře rozpoznatelných cílových stavů, kterých lze dosáhnout prostřednictvím omezeného souboru nástrojů; příkladem takové úlohy je Hanojská věž nebo lékařská diagnostika; jedná se také o úlohy, jejichž řešení se zdá lidem obtížné a namáhavé – alespoň tedy ve srovnání s percepčními a senzomotorickými úlohami, které lidé řeší bez jakékoli zjevné námahy, rychle, efektivně a u kterých si ani neuvědomují, že vlastně nějakou úlohu řeší (mezi takové úlohy by patřily například chůze, orientace v prostoru, chytání míče, vkládání nádobí do myčky, řízení auta, rozpoznávání tváří lidí nebo schopnost číst texty napsané různým typem písma).

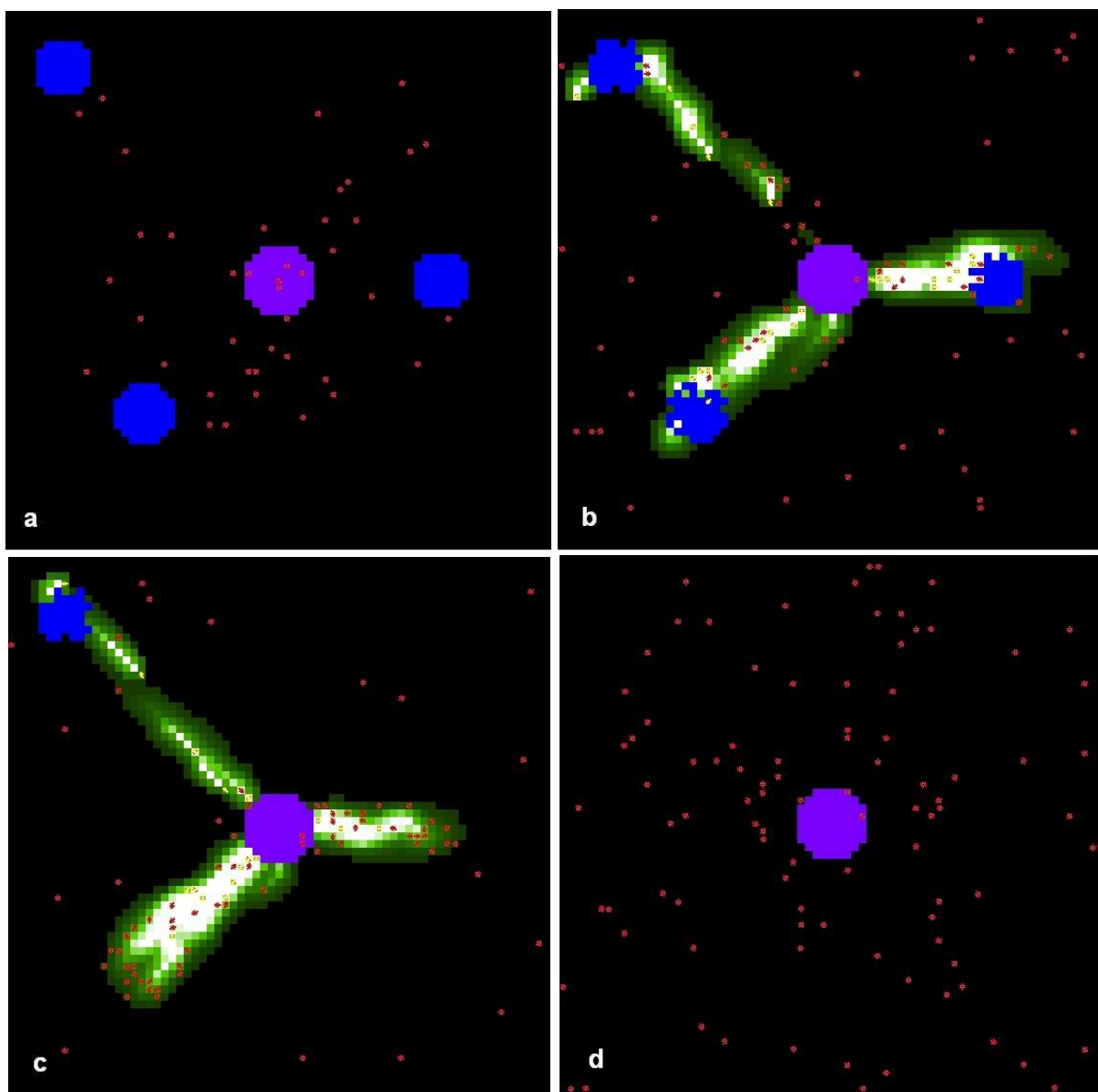
¹ „...generating and progressively modifying symbol structures until it reaches a solution structure.“

² „Everything of interest in cognition happens above the 100 millisecond level – the time it takes you to recognize your mother.“

1.2.3.3 Konekcionismus

Poněkud odlišný „výpočetní“ příběh než teorie fyzikálního symbolového systému (dále TFSS) nabízí konekcionismus. Konekcionismus stejně jako TFSS předpokládá, že lidská mysl je formou výpočtu. To, v čem se mezi sebou tyto dva přístupy liší, je názor na povahu tohoto výpočtu, resp. na povahu výpočetní architektury, která tento výpočet provádí. Zatímco TFSS vychází z předpokladu, že mentální výpočetní operace jsou definovány na symbolických strukturách, které lze relativně přímočaře interpretovat jako reprezentace běžných pojmů mentalistického diskurzu, konekcionistické výpočetní operace jsou definovány na objektech, které tak těsně se strukturou mentalistického diskurzu nekorespondují. Jinými slovy, TFSS předpokládá výpočetní operace na symbolové úrovni, zatímco konekcionismus na úrovni subsymbolové. K pochopení příčin a důsledků tohoto rozdílu mezi symbolovými a konekcionistickými systémy je potřeba mít alespoň základní představu o tom, na základě jakých principů konekcionistické systémy fungují.

Základní myšlenka konekcionismu je na určité úrovni abstrakce totožná s konceptem tzv. **decentralizovaných a distribuovaných komplexních systémů**. Ty *Melanie Mitchellová* (2006, s. 2) definuje jako rozsáhlé sítě velkého množství relativně jednoduchých agentů, jejichž lokální a omezenou – a z žádného privilegovaného centra nekoordinovanou – interakcí vzniká (emerguje) komplexní, koordinované a adaptivní chování. „Relativně jednoduché agenty“ v definici vyjadřují tu skutečnost, že ve srovnání s celkovým chováním systému je funkční role jeho jednotlivých prvků relativně jednoduchá - jakkoli sama o sobě může být dosti komplexní a sofistikovaná. „Emergentní komplexní a adaptivní chování“ pak odkazuje k chování, které (vedle toho, že je složité a že slouží k dosažení určitého cíle) povstává z kolektivní a paralelní činnosti většího množství jednoduchých prvků a přitom ho nelze přímočaře odvodit z jejich individuální činnosti (jinak řečeno, celek je vždy více než pouhý souhrn svých částí). Často citovanými příklady komplexních systémů jsou imunitní systém, biologické buňky, metabolické sítě, Internet, World Wide Web, ekonomické trhy, sociální sítě či kolonie sociálního hmyzu jako jsou mravenci nebo termiti (Barabási, 2005). (Na příloženém CD je v adresáři *StarLogo* obsažen instalační soubor programu StarLogo, který představuje uživatelsky vstřícné počítačové prostředí pro simulaci multiagentových systémů, které zainteresovanému čtenáři umožní prostřednictvím řady ukázkových simulačních projektů (nebo i prostřednictvím vlastních simulací) získat zkušenost s často protiintuitivním, překvapivě komplexním a dobře koordinovaným chováním multiagentových systémů sestávajících z velkého množství jednoduchých, tzv. reaktivních



Obrázek 35: Počítačová simulace hledání potravy kolonií mravenců. Na obrázcích (a, b, c d) je zachycen vývoj jednoho z ukázkových simulačních projektů, který je obsažen v simulačním prostředí StarLogo pod názvem „ants“. V této simulaci je modelováno hledání potravy kolonií 100 mravenců. Přestože každý jednotlivý mravenec (červené pixely) se řídí sadou velice jednoduchých pravidel, kolonie mravenců jako celek vykazuje překvapivě komplexní chování v podobě koordinované exploatace potravinových zdrojů nacházejících se v jeho okolí. Když mravenec najde potravu (modré pixely), odnese ji zpátky do mraveniště (fialové pixely) a při tom za sebou zanechává chemickou stopu (tmavě zelené až bílé pixely), která se s určitou intenzitou postupně vypařuje. Když jiný mravenec na tuto chemickou stopu narazí, sleduje ji směrem ke zdroji potravy. Čím více mravenců takto využívá určitý zdroj potravy, tím více se posiluje chemická stopa signalizující existenci a umístění tohoto zdroje potravy. Tato pozitivní zpětná vazba zaručuje koordinovaný způsob exploatace potravinových zdrojů v okolí mraveniště.

agentů, jejichž individuální činnost se řídí pouze několika málo jednoduchými pravidly specifikujícími omezené způsoby jejich reakce na různé struktury vyskytující se (nejčastěji v důsledku činnosti dalších agentů) v jejich prostředí.¹

Podle konekcionistů je podobným komplexním systémem také lidská mysl. *David Rumelhart* a kol. (1986b, s. ix) takto tvrdí, že „*intelligence povstává z interakce velkého počtu jednoduchých procesních jednotek*“². To je ale tak obecné konstatování, že se pod něj vejde řada dalších ne-konekcionistických výpočetních modelů (jako je například Brooksova subsumpční architektura, Minskyho teorie společenství mysli nebo Hofstadterův a Mitchellové *Copycat*³). *Dan Lloyd* (1989, s. 90) trochu blíže specifikuje povahu oněch jednoduchých procesních jednotek, když tvrdí, že „*základní myšlenkou konekcionismu je to, že kognice může být modelována simultánní interakcí velkého množství vzájemně vysoce propojených, neuronům-podobných jednotek*“⁴. Tzn. že konekcionismus se při modelování kognitivních procesů inspiruje (ovšem pouze do určité míry) lokální strukturou a funkcí biologického nervového systému, který je ukázkovým příkladem komplexního systému, jehož jednotlivé funkční komponenty (neurony) plní funkce, které jsou nesrovnatelně jednodušší než funkce nervového systému jako celku (rozpoznávání objektů, orientace v prostoru, senzomotorika, analogické myšlení, řešení problémů, hraní šachu, produkce a porozumění řeči atd.).

Konekcionistické modely - také umělé neuronové sítě (*ANN, artificial neural networks*) nebo systémy paralelního zpracovávání informací (*PDP, parallel distributed processing*) – se skládají - podobně jako biologický nervový systém - ze dvou základních komponent: z jednoduchých procesorů a z vážených (různě efektivních) spojení mezi těmito procesory.

¹ Na příloženém CD je v adresáři *StarLogo* kromě instalačního souboru obsaženo také několik html souborů s bližšími informacemi o způsobu ovládání simulačního prostředí *StarLogo*. Všechny tyto soubory jsou jinak volně dostupné z webových stránek <http://education.mit.edu/starlogo/>.

² „... *intelligence emerges from the interactions of large numbers of simple processing units.*“

³ *Copycat* je výpočetní model mentálních mechanismů, které leží v základech schopnosti analogického myšlení a vysokoúrovňové percepce (tj. té fáze zpracovávání smyslových podnětů, ve které hrají zásadní roli pojmy). Cílem Hofstadtera a Mitchellové při tvorbě *Copycatu* bylo dokázat, že jeden ze základních aspektů lidské kognice lze modelovat jako druh decentralizovaného a distribuovaného komplexního systému. Na příloženém CD lze v adresáři *Metacat* a pod souborem *Metacat-1.0* nalézt zdrojový kód k počítačovému programu *Metacat* od J. B. Marshalla, který představuje upravenou verzi *Copycatu* rozšířenou o schopnost „sebereflexe“ vlastního postupu řešení jednotlivých analogií (které - stejně jako v případě *Copycatu* - mají podobu proporcionálních analogií sestávajících z posloupnosti písmen abecedy, např. „Jestliže se *abc* mění na *abd*, jak se změní *xyz*?“). V html souboru *Metacat Home Page* lze nalézt všechny potřebné informace k instalaci a ovládání programu *Metacat*. Příloženy jsou rovněž tři práce od J. B. Marshalla a Melanie Mitchellové, které podrobněji popisují principy, na základě kterých oba modely pracují. Všechny zde uvedené zdroje jsou jinak volně dostupné z webových stránek projektu *Metacat* (<http://www.cs.pomona.edu/~marshall/metacat/>) a z osobních webových stránek Melanie Mitchellové (<http://web.cecs.pdx.edu/~mm/>).

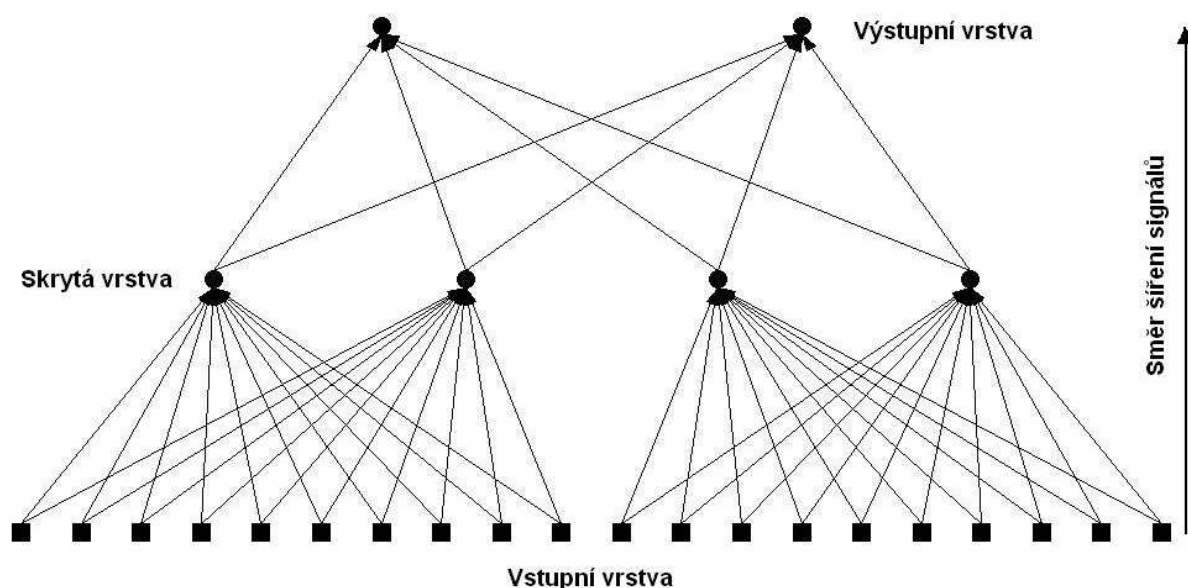
⁴ „*The central idea of connectionism is that cognition can be modeled as the simultaneous interaction of many highly interconnected neuronlike units.*“

V biologickém nervovém systému se jednoduché procesory nazývají **neurony** a vážené spoje mezi nimi pak **axony** a **synapse**. V konekcionistických modelech se tyto základní komponenty nazývají **jednotky** (*units*) či **uzly** (*nodes*), resp. **spoje** (*connections*). Přestože je konekcionistický model tvořen sítí hustě propojených uzlů, každý jednotlivý uzel přijímá vstupní informace pouze z relativně omezeného počtu¹ sousedních uzlů. Každý uzel na základě sumace takto přijatých vstupních informací a na základě určité jednoduché (většinou nelineární) matematické (tzv. **aktivační** nebo také **převodní**) **funkce** následně vypočítá výstupní informaci o **míře své aktivace**, kterou odešle prostřednictvím váženého spoje k uzlům, se kterými je náležitým způsobem propojen. Takto odeslaná informace pak slouží jako vstup do výpočetní procedury dalších neuronů, které počítají svou vlastní míru aktivace na základě míry aktivace sousedních uzlů, se kterými jsou propojeny prostřednictvím vážených spojů. Váhy těchto spojů přitom mohou být buď **pozitivní (excitační)**, nebo **negativní (inhibiční)**. Signál, který uzel takto na svém vstupu přijímá, je tedy vždy funkcí míry aktivace sousedních uzlů a vah spojů, po kterých se k němu signály o míře aktivace sousedních uzlů šíří. Důležité je, že jednotlivé uzly sítě mohou svou práci (sumaci vstupních signálů, výpočet míry aktivace a odeslání signálu k dalším uzlům) vykonávat do značné míry nezávisle jeden na druhém; díky tomu mohou operovat paralelně velké skupiny jednotek (proto *systémy paralelně distribuovaného zpracovávání informací*).

V konekcionistickém modelu lze podobně jako v biologickém nervovém systému rozlišit – řečeno velmi zjednodušeně - tři základní typy procesorů: jedny fungují jako druh receptorů, které přijímají své vstupy pouze z vnějšího prostředí; další uzly, které jsou jakousi obdobou interneuronů, jsou odříznuty od jakéhokoli kontaktu s vnějším světem a jsou jím ovlivňovány a ovlivňují ho vždy pouze prostřednictvím jiných uzlů; posledním typem uzlů jsou pak efekторы, které přijímají signály od jiných uzlů, ale své vlastní signály posílají již mimo vlastní neuronovou síť. Typickým uspořádáním těchto uzlů je tzv. **třívrstevná dopředná neuronová síť** (*three-layered feedforward neural network*), která se skládá ze tří vrstev neuronů (**vstupní**, **skryté** a **výstupní**) a ve které se informace o míře aktivace jednotlivých uzlů šíří vždy pouze jedním směrem, a to od vstupní vrstvy k vrstvě výstupní (viz obrázek

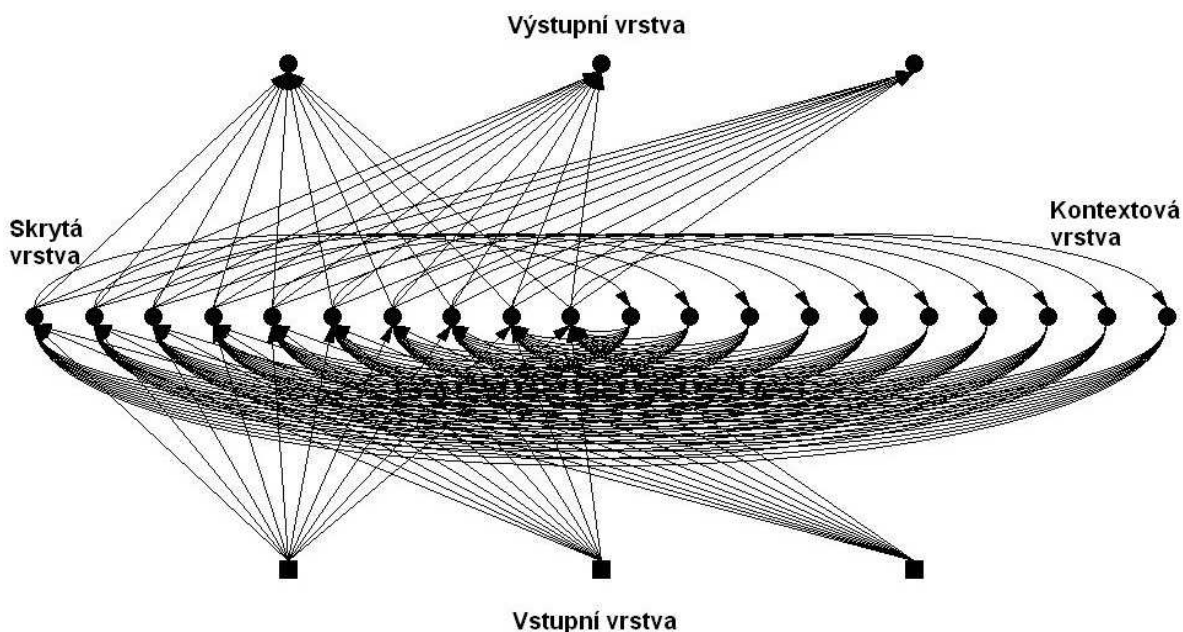
¹ Relativně omezenému vzhledem k celkovému počtu uzlů v síti. V mozku je takto jeden neuron propojen „pouze“ s $3-10 \times 10^3$ dalších neuronů, což je relativně zanedbatelné číslo, vezmeme-li v úvahu, že v mozku se (podle současných odhadů) nachází zhruba 10^{11} neuronů (Koukolík, 2005, s. 20): „Vzájemné zapojení mezi neurony, které jsou vzájemně vzdálené 1mm, je podstatně hustší než mezi neurony vzdálenějšími. [...] Pravděpodobnost přímého kontaktu mezi dvěma libovolnými neurony v jednom korovém sloupečku, jehož průměr se pohybuje přibližně od desetiny do poloviny milimetru, je asi 1:100. pravděpodobnost přímého kontaktu je pro vzájemně vzdálené neurony asi jedna ku milionu. Z toho plyne, že se v mozku na velké vzdálenosti předávají jen malé zlomky místní aktivity, a to prostřednictvím malého podílu místních neuronů.“ (Koukolík, 2006, s. 12-13)

36). Jiným typem architektury neuronové sítě s poněkud odlišným vzorcem šíření signálů je pak například tzv. **jednoduchá rekurentní síť** (*simple recurrent network*; viz obrázek 37).



Obrázek 36: Třívrstevná dopředná neuronová síť. Klasické uspořádání uzlů neuronové sítě do vstupní, skryté a výstupní vrstvy (input/hidden/output layer). (vytvořeno s pomocí simulačního softwaru Tlearn)

Zpracovávání informací v neuronové síti probíhá prostřednictvím šíření signálů (o míře aktivace jednotlivých uzlů) po vážených spojích mezi jednotlivými uzly. Správným způsobem modulovaný tok signálů v neuronové síti s tou správnou architekturou (tj. se správným počtem uzlů a s jejich správným zapojením) je přitom schopen implementovat jakoukoli vstupně-výstupní funkci. Tok signálů v neuronové síti – vedle její architektury - modulují především váhy jednotlivých spojů mezi uzly sítě. Váhy spojů fungují jako jakási stavidla, která požadovaným způsobem omezují a regulují tok mediační látky, která nese informaci o míře aktivace jednotlivých uzlů sítě. Tok mediační látky neuronovou sítí je takto „vytvarován“ do podoby, kterou mu diktují různá omezení v podobě vah spojů mezi uzly sítě. Tento způsob zpracovávání informací, který stojí v pozadí činnosti neuronových sítí, se obecně nazývá **paralelní respektování omezení** (*parallel constraint satisfaction*), jehož „jednoduchou ilustrací...je práce fakultních rozvrhářů na začátku semestru. Nemohou současně rozvrhovat do jedné místnosti víc kurzů, a ani studenti a jejich profesori nemohou být současně na více místech. K tomu přistupují specifické požadavky profesorů i studentů týkající se doby a místa konání jednotlivých kurzů.“ (Thagard, 2001, s. 129)



Obrázek 37: Jednoduchá rekurentní síť. Na rozdíl od klasické dopředné neuronové sítě, ve které se signály mohou šířit pouze jedním směrem a to od uzlů ve vstupní vrstvě k uzlům ve výstupní vrstvě, u jednoduché rekurentní sítě mohou existovat spoje i v rámci jednotlivých vrstev. Tzv. kontextová vrstva slouží jako paměť, do které se kopíruje aktuální míra aktivace jednotlivých uzlů ve skryté vrstvě. Tuto informaci uzly v kontextové vrstvě v následujícím aktualizacím cyklu (kdy uzly v jednotlivých vrstvách počítají svou míru aktivace na základě přijatých signálů) odesílají zpátky k uzlům ve skryté vrstvě, které tak svoji míru aktivace počítají nejen na základě signálů přicházejících ze vstupní vrstvy, ale také na základě svého vlastního předchozího stavu (tj. na základě své míry aktivace z předchozího aktualizacího cyklu). Tento způsob zapojení umožňuje jednoduché rekurentní síti implementovat funkce, které nezávisí pouze na aktuálním vstupu, ale také na dřívější historii vstupů. (vytvořeno s pomocí simulačního softwaru Tlearn)

Z výše uvedeného vyplývá, že k implementaci požadované funkce je vždy potřeba nalézt vhodným způsobem nastavené váhy spojů. Vzhledem k obrovskému počtu spojů v neuronových sítích je prakticky nemožné nastavovat jejich váhy „ručně“; proto se k nalezení váhové matice, která by dokázala implementovat požadovanou funkci, využívají různé biologicky více či méně (ale spíše méně) plausibilní automatizované mechanismy – tzv. **učící** či **tréninkové algoritmy**, které jsou obdobou vyhledávacích heuristik, které lidé využívají při prohledávání stavového prostoru nějaké úlohy. V případě učících algoritmů neuronových sítí je prohledávaným stavovým prostorem množina všech možných váhových matic dané neuronové sítě a cílem učícího algoritmu je pak v tomto pomyslném prostoru najít takovou váhovou matici, která by dokázala usměrnit tok signálů mezi uzly sítě způsobem umožňujícím požadovanou transformaci vstupů na výstupy. Schopnost neuronové sítě produkovat na základě vstupních informací odpovídající výstup takto není výsledkem

pečlivého naprogramování nějakého lidského programátora, ale spíše výsledkem její vlastní schopnosti se učit, tedy najít tu správnou váhovou matici, ve které jsou implicitně zakódovány všechny znalosti, které potřebuje k tomu, aby dokázala vyřešit daný problém.

Váhová matice spolu s architekturou neuronové sítě takto funguje jako „dlouhodobá paměť“ či „znalostní báze“, ve které jsou uloženy všechny znalosti potřebné k řešení vybraných úloh. Uloženy jsou přitom dost netradičním způsobem. Zatímco v případě symbolového systému (jako je například produkční systém) jsou všechny potřebné znalosti zakódovány explicitně v podobě různých symbolických reprezentací, u neuronové sítě nic takového nenajdeme. Vše potřebné je zde implicitně zakódováno do mnoha vah spojujících uzly a do celkové architektury sítě. Takto kódované znalosti (tj. jednotlivé vážené spoje) navíc často participují na řadě různých úloh, které neuronová síť jako celek řeší. Podobně distribuovanou povahu má také obdoba krátkodobé „pracovní paměti“ neuronové sítě, která kóduje aktuálně využívané informace při zpracovávání určitého specifického vstupu prostřednictvím přechodných vzorců aktivace většího množství uzlů (zejména těch ve skryté vrstvě). Tímto jsme se dostali k podstatě hlavního rozdílu mezi symbolovými systémy a konekcionistickými modely, který spočívá v odlišném způsobu reprezentace informací v těchto dvou typech výpočetních systémů.

Konekcionistické modely využívají tzv. **distribuovaných** či **superpozičních reprezentačních schémat**, tzn. že k reprezentaci nějaké informace používají souběžnou aktivitu většího množství uzlů, které přitom mohou participovat na reprezentaci většího množství různých informací. K identifikaci takto reprezentované informace potom tedy nestačí znát míru aktivace pouze jednoho uzlu sítě, ale vždy celý vzorec aktivace určité skupiny uzlů. Neuronové sítě dokáží využívat tohoto distribuovaného způsobu kódování k reprezentaci informací o velice jemných, mikrostrukturálních rozdílech a podobnostech mezi různými pojmy, předměty, událostmi či situacemi. Určitý vzorec aktivace několika uzlů sítě takto například může reprezentovat nějaký objekt, který se právě vyskytuje v zorném poli a který má určitou specifickou orientaci; nepatrnou změnou tohoto vzorce aktivity, například změnou míry aktivace jednoho či dvou uzlů sítě, lze pak kódovat informace o změnách orientace tohoto objektu. Přestože oba dva aktivační vzorce se do značné míry překrývají, nepatrné rozdíly, které mezi nimi existují, nejsou bezvýznamné, neboť nesou informaci o důležitých rozdílech v orientaci daného objektu. Distribuovanou reprezentaci lze takto využít také ke kódování některých důležitých podobností mezi různými objekty. Například aktivační vzorec reprezentující šimpanze by mohl sdílet velkou část aktivačního vzorce reprezentujícího gorilu, ale jen velice nepatrnou část aktivačního vzorce kódujícího žraloka. Právě toto

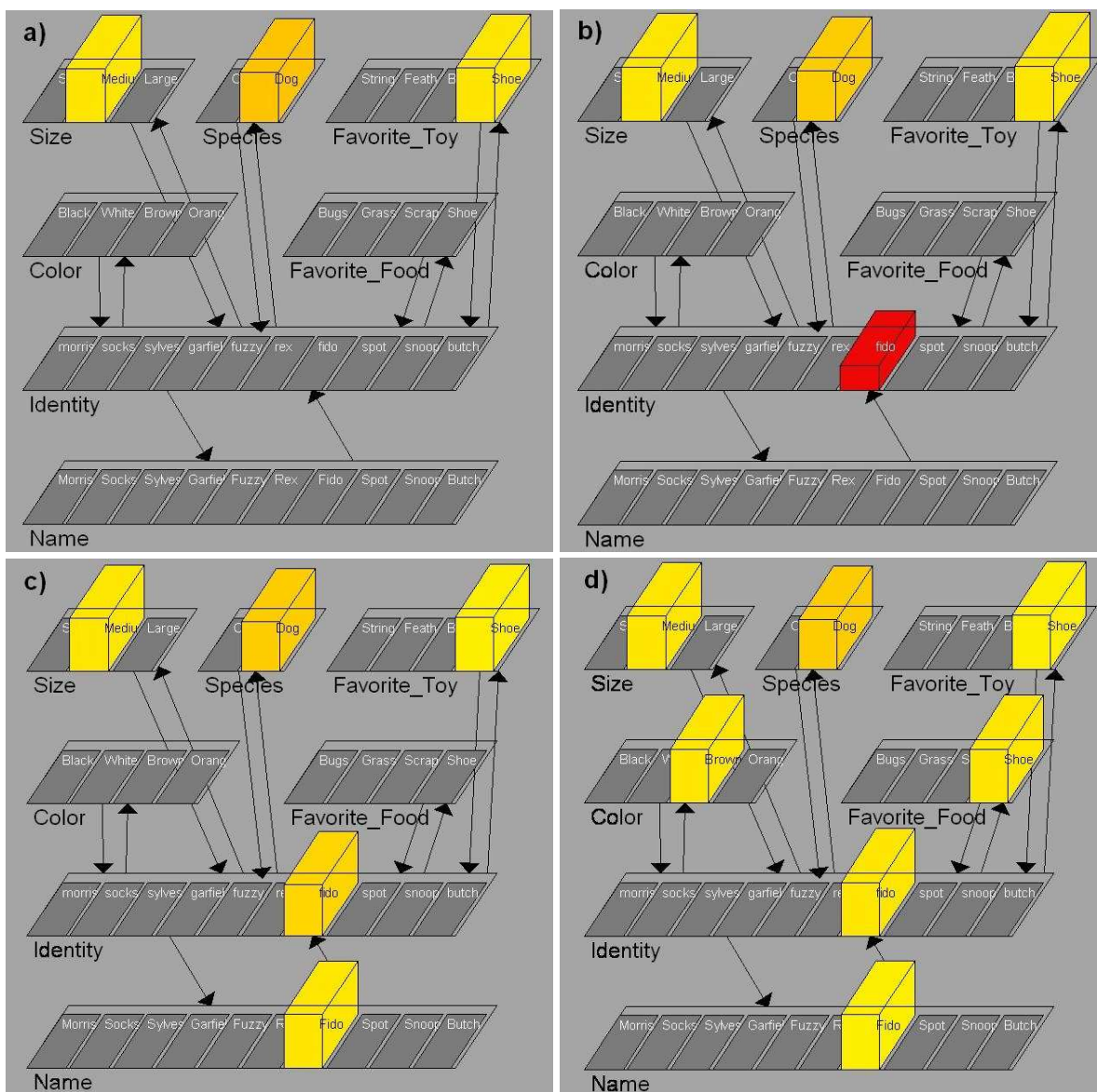
překrývání a sdílení reprezentačních zdrojů zachycuje termín *superpoziční reprezentační schéma*. Z hlediska výpočtů prováděných umělou neuronovou sítí je důležité, že toto překrývání a sdílení reprezentačních zdrojů je systematické a tudíž informačně signifikantní - tzn. že ne/sdílené reprezentační zdroje nám (resp. neuronové síti) říkají něco důležitého o ne/podobnosti různých objektů či událostí.

Tento způsob reprezentace informací má řadu žádoucích vedlejších efektů. Tím první je vysoce **efektivní využívání reprezentačních zdrojů**, které je důsledkem toho, že jednotlivé uzly sítě mohou participovat na reprezentaci většího množství různých objektů. Druhým vedlejším efektem je **schopnost generalizace**, tj. schopnost aplikovat nabyté znalosti i na objekty, se kterými se neuronová síť doposud neměla šanci potkat. Tuto schopnost má neuronová síť tzv. „zdarma“, neboť je automatickým důsledkem distribuovaného způsobu reprezentace informací: Jestliže se aktivační vzorec kódující nějaký nový a neznámý předmět alespoň částečně překrývá s aktivačním vzorcem nějakého jiného a známého předmětu, reakce neuronové sítě na nový předmět bude vycházet z tohoto částečného překrytí dvou aktivačních vzorců. Jedná se de facto o praktickou implementaci myšlenky, kterou formuloval již *David Hume* (1972, s. 68), totiž že „u podobných příčin čekáme podobné účinky“. Takto se například může neuronová síť prostřednictvím sady konkrétních příkladů (a určitého učicího algoritmu) naučit tomu, že šimpanzi mají rádi banány. Jestliže neuronová síť bude reprezentovat šimpanze prostřednictvím aktivačního vzorce, jehož velkou část bude sdílet také aktivační vzorec reprezentující gorilu, potom neuronová síť dokáže generalizovat svůj poznatek o šimpanzích i na gorily, a dojít tak tedy k závěru, že gorila má ráda banány, a to přesto, že neuronové síti nikdo nikdy explicitně neřekl, že by gorila měla ráda banány. To, že to neuronová síť dokáže, je čistě důsledkem toho, že oba aktivační vzorce navzájem sdílejí značnou část uzlů sítě; a to, že to je navíc pravda (že gorily opravdu mají rády banány), je pak důsledkem toho, že toto překrytí aktivačních vzorců kódujících dva různé objekty je systematické a nenáhodné a že odráží určité zákonitosti, na základě kterých svět funguje.

Dalším zajímavým vedlejším efektem distribuovaného reprezentačního schématu je schopnost neuronové sítě implementovat tzv. **obsahem-adresovanou paměť** (*content-addressable memory*), která k vybavení vzpomínek využívá různá vodítka, která jsou uložena v paměti v podobě řady dílčích informací, ze kterých se vybavovaná vzpomínka skládá nebo se kterými tato vzpomínka nějak jinak úzce souvisí. Tento způsob paměťového vybavování informací je charakteristický pro člověka, který když se snaží vzpomenout si na nějakou informaci, vždy už má nějakou představu o tom, co vlastně hledá. Většina tradičních externích úložišť informací - jako je například výkladový slovník nebo počítačová paměť -

funguje na poněkud odlišném principu. V těchto systémech má každé místo, na kterém je uložena nějaká informace, svoji adresu; a znalost této adresy je jediný způsob, jak se k takto uložené informaci dostat. Například ve výkladovém slovníku je adresou posloupnost písmen hledaného slova a definice slova, kterou lze na této adrese najít, je pak onou hledanou informací. Rozdíl mezi těmito dvěma odlišnými způsoby vybavování informací lze ilustrovat na tom, jak se k požadované informaci dostaneme s pomocí slovníku a jak s pomocí nějaké jiné osoby. Řekněme, že bychom chtěli znát název románské církevní stavby na centrálním kruhovém půdorysu s jednou nebo více absidami. U tradičního výkladového slovníku neexistuje způsob, jak zjistit místo, kde se požadovaná informace (ROTUNDA) nachází, neboť jediný způsob, jak lze se slovníkem pracovat, je že člověk nejdříve zjistí adresu (ROTUNDA), najde ve slovníku místo s touto adresou a nakonec si přečte informaci, která je na této adrese uložena. Bez hesla se ovšem člověk žádné užitečné informace nedobere. Naproti tomu osoba, které poskytneme alespoň část výše uvedených informací (tedy že se jedná o románskou církevní stavbu...), bude většinou schopna poměrně rychle zpřístupnit i zbytek souvisejících informací (a mezi nimi i název stavby). Lidská paměť si takto narozdíl od slovníku dokáže zjednat přístup k nějaké informaci prostřednictvím řady dalších informací, které jsou v paměti uloženy a které s hledanou informací nějak souvisí (McLeod, Plunkett, Rolls, 1998). Obsahem-adresovanou paměť umožňuje ta skutečnost, že informace jsou v paměti kódovány prostřednictvím distribuovaných reprezentací, takže i zpočátku jen částečná aktivace uzlů kódujících vzpomínku na daný předmět nebo událost vede nakonec k „rekonstrukci“ celého aktivačního vzorce, ve kterém je zakódována i hledaná dílčí informace, například o názvu románské církevní stavby.

Díky tomuto mechanismu je obsahem-adresovaná paměť také **odolná vůči chybným nebo neúplným vstupním informacím** (*fault toleration*). Například jestliže někdo bude chtít znát jméno nějaké osoby a tuto osobu bude popisovat jako člověka, který byl v roce 1999 zvolen 43. prezidentem USA, za jehož úřadování došlo k teroristickým útokům na USA a který nařídil vojenskou intervenci nejdříve do Afghánistánu a později do Iráku s cílem odstranit tamější režimy (údajně) ohrožující světovou bezpečnost, potom bude asi většina lidí schopná bez větších problémů odpovědět, že touto osobou je George W. Bush, a to přesto, že v jeho charakteristice se vyskytuje chyba (G. W. Bush se stal prezidentem až v roce 2000). Mechanismus stojící v pozadí této odolnosti vůči informačnímu šumu je totožný s tím, který neuronové síti umožňuje generalizovat na nové, dosud nepotkané případy: aktivační vzorce kódující bezchybný a chybný popis G. W. Bushe se do značné míry překrývají, takže v obou případech bude neuronová síť reagovat obdobným způsobem.



Obrázek 38: Obsahem-adresovaná paměť. Na obrázku jsou čtyři (a-d) „momentky“ zachycující postupné šíření aktivace mezi rekurentně propojenými vrstvami uzlů tvořících asociční paměť, která obsahuje informace o deseti různých domácích zvířatech (pěti psech a pěti kočkách), z nichž se každé vyznačuje určitou specifickou kombinací charakteristik (jméno, identita, barva srsti, velikost a oblíbené jídlo a hračka). Na obrázku je zachycen vývoj aktivačního vzorce v situaci odpovídající dotazu „Jak se jmenuje ten středně velký pes, co si rád hraje s botami?“ Ze série „momentek“ je patrné, jak si neuronová síť postupně vybavuje všechny informace, které má o zvířeti s popsány vlastnostmi k dispozici (tedy nejen to, že se jmenuje Fido, ale také to, že to je „právě ten“ pes Fido, že má hnědou srst a že se občas rád zakousne do bot). Psychologický proces vybavování informací z paměti je zde modelován jako proces postupného usazování se aktivačního vzorce uzlů sítě do té podoby, ve které nejlépe vyhovuje jednotlivým omezením v podobě dotazového aktivačního vzorce a v podobě vah spojů (mezi uzly sítě), ve kterých jsou zakódovány všechny znalosti o deseti různých domácích zvířatech, které si síť osvojila. (vytvořeno s pomocí simulačního softwaru PDP++ na základě cvičného příkladu z knihy Randalla O'Reillyho a Yuko Munakatové (2000); na přiloženém CD v adresáři „PDP“ je obsažen instalační soubor tohoto simulátoru a také manuál k jeho ovládání; software je jinak také volně k dispozici na webových stránkách <http://www.cnbc.cmu.edu/Resources/PDP++//PDP++.html>)

V této souvislosti se hovoří také o tzv. **požvolné degradaci** (*graceful degradation*) výkonu neuronové sítě v důsledku lokálního poškození nebo odstranění některých z jejích funkčních komponent (tj. uzlů a spojů mezi nimi). Zatímco tradiční symbolové systémy na poškození některé ze svých komponent většinou reagují okamžitou a naprostou ztrátou své funkce, konekcionistické systémy jsou k takovému katastrofickému selhání mnohem méně náchylné; jejich výkon s narůstajícím poškozením spíše pozvolna klesá a do určité kritické míry poškození jsou schopny relativně úspěšně vykonávat svou funkci, tedy zpracovávat přicházející informace. Takto poškozená neuronová síť funguje jako tzv. **kompletovač vzorců** (*pattern completer*), který je schopen i na základě pouze částečného vstupu zkompletovat původní vzorec či strukturu. Tato schopnost kompletovat neúplné vzory a struktury je dána tím, že všechny znalosti nutné ke zpracovávání informací jsou v neuronové síti vždy roz distribuovány mezi velké množství jednotlivých uzlů a vazeb, takže poškození určitého procenta těchto zdrojů ještě nemusí představovat fatální narušení funkce neuronové sítě jako celku.

Z hlediska rozdílů, které existují mezi způsobem, jakým modelují mentální procesy symbolové a konekcionistické systémy, jsou důležité zejména dvě vlastnosti distribuovaného a superpozičního reprezentačního schématu. Tou první vlastností je to, co *Paul Smolensky* (1988) nazývá **dimenzionálně posunuté reprezentace** (*dimension shifted representations*), a tou druhou je pak **jemná citlivost distribuovaných reprezentací na změnu kontextu**. Obě tyto vlastnosti jsou důsledkem té skutečnosti, že distribuované reprezentace jsou **subsymbolové**, tzn. **sémanticky netransparentní** (viz s. 69-71): Na rozdíl od symbolových systémů, jejichž vnitřní stavy relativně přímočaře odpovídají pojmům běžného diskurzu, konekcionistické systémy nemodelují mentální procesy prostřednictvím výpočetních manipulací vnitřních stavů, které by těsně korespondovaly s běžnými pojmy zachycujícími - introspekci více či méně dostupné - různé mentální stavy a procesy (přesvědčení, přání, cíle, schémata, vjemy apod.). To, v čem se oba systémy liší, je tedy to, zda jsou jimi definované výpočetní operace aplikovány na objekty, které lze přímočaře interpretovat v termínech mentalistického diskurzu: „V symbolovém přístupu se symboly (atomy) používají k označení sémanticky interpretovatelných entit (pojmu). Ty samé symboly jsou pak objektem manipulace podle pravidel, která definují systém. Entity, které lze sémanticky interpretovat, jsou tedy zároveň entitami, které podléhají formálním pravidlům, která definují systém. V rámci subsymbolového paradigmatu toto již neplatí. Sémanticky interpretovatelnými entitami jsou vzorce aktivace velkého počtu jednotek v systému, zatímco entitami manipulovanými podle formálních pravidel jsou jednotlivé míry aktivace uzlů sítě. Tato pravidla mají podobu

*pravidel pro výpočet míry aktivace, což jsou pravidla značně odlišná od pravidel pro manipulaci symbolů.*¹ (Smolensky, 1987, s. 100) Zatímco tedy symbolové systémy aplikují své výpočetní operace přímo na objekty, které se nacházejí blízko úrovně vědomého a introspekci nahlédnutelného myšlení, subsymbolové systémy jsou definovány výpočetními operacemi, které jsou aplikovány na entity, které jsou od úrovně vědomého myšlení poměrně vzdálené. To je také důvod, proč Smolensky popisuje subsymbolové systémy jako dimenzionálně posunuté. Míry aktivace jednotlivých uzlů sítě většinou nenesou stejný typ významu jako slova běžného jazyka. To ale neznamená, že by nenesly význam vůbec žádný. Pouze se jedná o významy tak jemné a subtilní, že je lze jen velice obtížně zachytit prostřednictvím pojmů běžného diskurzu. Změny v úrovni aktivace jednoho uzlu takto mohou kódovat informace o různých, velice jemných kontextuálních odstínech právě zpracovávaného podnětu. V běžných, vysokoúrovňových pojmech lze interpretovat pouze aktivační vzorce větší skupiny uzlů sítě.

V této souvislosti si pak lze klást otázku, zda by nebylo možné považovat tyto rozsáhlé aktivační vzorce za ony koreláty vysokoúrovňových pojmů, které by byly podrobovány výpočetním manipulacím v souladu s pravidly definujícími konekcionistický systém. Nebo ještě jinak řečeno: *Nejsou konekcionistické systémy přece jenom na určité úrovni popisu sémanticky transparentní?* Odpověď na tuto otázku zní: *Nejsou*. Proč tomu tak je, lze ukázat na velice jednoduchém příkladě: Řekněme, že budeme mít větu „pes honil kočku po zahradě“. Běžný počítačový (symbolový) systém, který by se tuto větu snažil přeložit například do angličtiny, by zcela jistě disponoval určitými vnitřními stavy (v nějakém programovacím jazyce a na té nejnižší úrovni pak ve strojovém kódu), které by označovaly slova „pes“ a „kočka“. V případě konekcionistického systému by při zpracovávání výše uvedené věty byla aktivní řada různých uzlů sítě. S vysokou pravděpodobností by bylo možné izolovat určité specifické vzorce aktivace, které by odpovídaly slovům „pes“ a „kočka“. Co je ale důležité je to, že tyto aktivační vzorce by nebyly – na rozdíl od symbolických struktur tradičního systému – stabilní napříč různými kontexty, ve kterých se slova „pes“ a „kočka“ mohou vyskytovat. Například slovu „pes“ ve větě „pes honil kočku po zahradě“ by odpovídal aktivační vzorec, který by byl mírně odlišný od aktivačního vzorce odpovídajícího slovu

¹ „In the symbolic approach, symbols (atoms) are used to denote the semantically interpretable entities (concepts). These same symbols are the objects governed by symbol manipulations in the rules which define the system. The entities which are capable of being semantically interpreted are also the entities governed by the formal laws that define the system. In the subsymbolic paradigm, this is no longer true. The semantically interpreted entities are patterns of activation over a large number of units in the system, whereas the entities manipulated by formal rules are the individual activations of cells in the network. The rules take the form of activation passing rules, of essentially different character from symbol manipulation rules.“

„pes“ ve větě „pes ležel ve svém kutlochu na zahradě“. V případě první věty by aktivační vzorec mohl zahrnovat aktivní jednotky kódující například agresivitu, násilí nebo ostré zuby; zatímco v případě druhé věty by tyto jednotky aktivní nebyly (aktivní by byla například jednotka kódující pocit pohody). Přestože se tedy oba aktivační vzorce budou do značné míry překrývat (protože v obou scénách bude vystupovat pes), nebudou zcela totožné, takže nebude možné o nich hovořit jako o přímočarých korelátech vysokoúrovňového pojmu pes. A právě tímto se subsymbolový systém liší od symbolového systému, v němž je slovo „pes“ - bez ohledu na kontext - reprezentováno stále stejně, takže jeho vnitřní syntaktický korelát je možné bez problémů propojit s vysokoúrovňovým pojmem psa. Smolensky (1988, s. 17) tento rozdíl mezi oběma typy výpočetních systémů komentuje následujícími slovy: „*V symbolovém paradigmatu se kontext symbolu manifestuje v jeho okolí a skládá se z dalších symbolů; v subsymbolovém paradigmatu se kontext symbolu manifestuje uvnitř něho samotného a skládá se ze subsymbolů.*“¹ A právě tato skutečnost podle Clarka (1989) zakládá sémantickou netransparentnost konekcionistických systémů.

1.2.3.4 Výpočetní architektura mysli/mozku

Současnou kognitivní vědu charakterizuje určitá „schizofrenie“ ohledně názoru na to, jaká je základní povaha výpočetní architektury, která provádí mentální výpočet. Tato rozpolcenost má přitom svůj zdroj již v samotné definici kognitivní vědy, která bývá tradičně vymezována jako vědecká disciplína, jejímž předmětem zájmu je mysl a mozek. Problém tkví v tom, že mozek (jakožto biologický substrát lidského myšlení) a mysl (jakožto na biologickém substrátu sice závislá, ale spíše abstraktní a hůře postižitelná entita) se jeví být dvěma naprosto odlišnými výpočetními stroji. Mozek se takto podle všech dosavadních zjištění jeví být masivně paralelním počítačem, který se skládá z miliard jednouchých procesorů (neuronů), které manipulují s jednoduchými čísly v podobě míry aktivace jednotlivých neuronů a jejichž vzájemná interakce (účinnost synaptických spojů mezi neurony) se mění v souladu se statistickými charakteristikami prostředí, s nímž počítač interaguje. Na určité úrovni abstrakce jsou odpovídajícím modelem těchto neurálních výpočtů konekcionistické (či PDP) systémy. Lidská mysl a zejména její vyšší kognitivní funkce jako je jazyk, abstraktní myšlení, plánování, řešení problémů, analogické myšlení apod. se však zdají být produktem poněkud odlišné výpočetní architektury. Již v antice lze nalézt zárodky představy, že lidská

¹ „*In the symbolic paradigm the context of a symbol is manifest around it and consists of other symbols; in the subsymbolic paradigm the context of a symbol is manifest inside it, and consists of subsymbols.*“

mysl je druhem sériového procesoru, který manipuluje komplexními symbolickými strukturami na základě pravidel, která se podobají pravidlům formální logiky.¹ Tuto myšlenku později precizovali filozofové a vědci jako byli Descartes, Locke, Kant, Boole, Frege, Tarski či Turing, který ji nakonec (spolu s von Neumannem) převedl do podoby konkrétního návrhu výpočetní architektury moderního digitálního počítače, který je schopen s několika málo základními operacemi, s konečným množstvím arbitrárních symbolů a na základě těch správných instrukcí provést jakýkoli dostatečně dobře specifikovaný výpočet. Právě tato schopnost generovat neomezené množství reakcí s pomocí omezeného souboru prostředků se zdá být charakteristická pro řadu oblastí, kde se uplatňují vyšší mentální funkce. Klasickým příkladem je v tomto ohledu jazyková kompetence, která člověku umožňuje generovat neomezené množství smysluplných (i nesmyslných) jazykových výpovědí na základě omezeného souboru gramatických pravidel, která určují, jak se mohou slova kombinovat do frází a do vět podle jejich příslušnosti k jednotlivým slovním druhům. Jazykové výpovědi takto podle všech dosavadních zjištění (Pinker, 2003, 1999, 1994) tvoří komplexní hierarchické syntaktické struktury, které člověk prostřednictvím konečného souboru pravidel kóduje do lineárního sledu hlásek a písmen a které člověk z této lineární sekvence hlásek a písmen prostřednictvím stejných pravidel zase dekóduje. Lidská mysl se tedy ve svých výlučných kognitivních doménách (jako je právě jazyková kompetence nebo schopnost plánovat či řešit různé abstraktní problémy) jeví být určitým druhem kombinatorického výpočetního systému, který se skládá ze souboru systematicky kombinovatelných symbolických reprezentací a ze souboru procedur, které kromě běžné manipulace umožňují také dekompozici komplexních symbolických struktur na symbolové atomy či naopak jejich kombinaci do složitějších symbolových molekul. Modelem tohoto pojetí lidské mysli je tradiční symbolový systém (Fodor, Pylyshyn, 1988). Máme zde tedy dva různé typy modelů kognitivních procesů, u kterých se zdá, že spolu navzájem nijak nesouvisí. Vyvstává tak otázka, co tyto dva různé typy modelů vlastně modelují, jaký je mezi nimi vztah a co nám to říká o povaze lidského myšlení.

¹ S tímto pojetím lidské mysli se lze v té či oné podobě setkat téměř v každé středoškolské nebo vysokoškolské učebnici obecné psychologie. Například Rosina a kol. (1963, s. 98-103) ve své učebnici obecné psychologie pro pedagogické instituty popisují lidské myšlení jako proces, který probíhá ve formě *pojmu*, *soudů* a *úsudků*, kde 1) pojem představuje „*odraz obecných neboli podstatných vlastností předmětů a jevů v našem vědomí*“, který se konstituuje na základě vjemů a představ konkrétních objektů a událostí, 2) soud je vyjádřením „*pochopení vztahů [a souvislostí] mezi pojmy*“, a 3) úsudek je (deduktivním, induktivním nebo analogickým) „*vyvozením nového soudu [označovaného jako závěr] z jiných známých soudů [označovaných jako premisy] nebo... pochopením vztahů [a souvislostí] mezi soudy*“. Dítě si takto například může ve škole na základě řady konkrétních příkladů a výkladu učitele osvojit pojmy *kovu* a *vodiče* a poznatek o vztahu mezi těmito dvěma pojmy v podobě soudu *Kovy jsou vodiče*; dozví-li se pak dítě, že železo patří mezi kovy, může na základě deduktivního úsudku velice snadno dojít ke správnému závěru, že železo je vodičem elektrického proudu.

Oba dva modely (konekcionalistický i symbolový) jsou modely výpočetními. To, v čem se liší, je názor na povahu „jazyka myšlení“ či symbolického média, ve kterém se mentální výpočet odehrává. Rozdíl je tedy v úrovni, na které předpokládají, že probíhají výpočetní operace vysvětlující mentální procesy: Konekcionalistické modely – na rozdíl o symbolových modelů - předpokládají výpočetní operace na úrovni, která má mnohem blíže k neurovědě než k psychologii: To, co se v rámci konekcionalistického systému počítá, nejsou implikace nějaké obdoby logických výroků, které lze přímočaře interpretovat v pojmech běžného diskurzu, ale míry aktivace jednotek, které samy o sobě nenesou žádný význam, který by těsně korespondoval s vnějším světem předmětů a událostí: Sémantika symbolického média konekcionalistických systémů je fakticky omezena pouze na objekty a události, které se odehrávají uvnitř samotného výpočetního systému; aktivita jednotlivých jednotek konekcionalistického systému neoznačuje nic jiného než informace o míře aktivity sousedních jednotek, se kterými jsou propojeny. Symbolické médium konekcionalistických systémů tak lze přirovnat ke strojovému kódu počítače – základnímu programovacímu jazyku, ve kterém je možné bezprostředně vyjádřit pouze jednoduché instrukce týkající se toho, co a jak udělat s obsahem jakého paměťového registru, ale nic o vnějším světě předmětů a událostí. Jakkoli je podle všech dosavadních zjištění téměř jisté, že obdobné výpočetní procesy se v lidském mozku opravdu odehrávají, subjektivní zkušenost i řada výzkumů prováděných v rámci kognitivní vědy napovídají, že zde musí existovat určitý „psychologický virtuální stroj“, tj. určitá úroveň popisu, na které lze jednotlivým fyzikálním stavům konekcionalistického systému připsat významy běžných pojmů mentalistického diskurzu. To, v čem se pak kognitivní vědci liší, je názor na to, zda tato vyšší úroveň popisu má nebo nemá sama o sobě výpočetní charakter, tedy zda je možné chování kognitivního systému na této úrovni plně a beze zbytku zachytit prostřednictvím souboru formálních pravidel definovaných na objektech, které lze na této úrovni abstrakce identifikovat.

V této souvislosti se často hovoří o tzv. **eliminativistech** a **implementacionalistech** (Smolensky, Legendre, 2006): Podle eliminativistů „psychologický virtuální stroj“ výpočetní povahu nemá. Eliminativisté sice uznávají, že na určité úrovni popisu lidského kognitivního systému lze identifikovat entity, které těsně korespondují s pojmy mentalistického diskurzu, ale zároveň tvrdí, že vztahy a interakce mezi těmito entitami nemají výpočetní, ale pouze statistickou, emergentní či epifenomenální povahu. To potom znamená, že „psychologický virtuální stroj“ nelze formálně specifikovat (za použití jazyka odpovídajícího mentalistickému diskurzu) jako výpočet nějakého vysokoúrovňového algoritmu, který je implementován prostřednictvím výpočetní architektury lidského mozku. Podle eliminativistů se tak vztah

mezi mozkiem a myslí zásadním způsobem odlišuje od obvyklého vztahu mezi hardwarem a softwarem, v jejichž případě zde vždy existuje přímočarý překlad (implementace) vysokoúrovňového popisu funkce virtuálního stroje softwaru do strojového kódu hardwaru a vice versa: „V normálním programu můžete vysvětlit každou jednotlivou operaci na bitové úrovni tím, že se podíváte nahoru směrem k nejvyšší programové úrovni. Můžete sledovat, jak vysokoúrovňová funkce volá subrutiny, které volají další subrutiny, které volají určitou proceduru na úrovni strojového jazyka...[Vždy] zde existuje nějaký vysokoúrovňový, globální důvod toho, proč se s určitým konkrétním bitem [právě] manipuluje.“¹ (Hofstadter, 1985, s. 653) Toto podle eliminativistů pro vztah neurálního výpočetního zařízení (mozku) a „psychologického virtuálního stroje“ (mysli) neplatí: Výpočetní činnost jednotlivých funkčních jednotek konekcionistického systému není překladem nějakého vysokoúrovňového „mentálního programu“ do strojového kódu biologického hardwaru lidského mozku. V tomto ohledu se tedy – podle eliminativistů - symbolické médium konekcionistických systémů zásadním způsobem liší od běžného strojového kódu digitálního počítače. Důsledkem je rozchod se základním předpokladem počítačové vědy, totiž že konkrétní specifické detaily implementace nějakého algoritmu (softwaru, virtuálního stroje) nemění nic na podstatě realizovaného výpočtu, který je plně specifikován na vyšší úrovni než je úroveň základních operací daného hardwaru. Tento předpoklad člověku do značné míry² dovoluje ignorovat specifikum hardwaru, na němž je daný software implementován, a věnovat se pouze vyšší úrovni popisu realizovaného výpočtu - ve stylu programátora, který ignoruje úroveň strojového jazyka počítače a zabývá se pouze tím, jak zkombinovat výrazy nějakého vysokoúrovňového programovacího jazyka způsobem, který by mu umožnil realizovat požadovaný výpočet.

Z tohoto základního předpokladu počítačové vědy naopak vycházejí **implementacionalisté**, podle nichž je možné „psychologický virtuální stroj“ formálně specifikovat jako vysokoúrovňový výpočetní algoritmus (symbolický procesor), který je – pouze shodou historických náhod a okolností, souhrnně nazývaných „evoluce přírodním výběrem“ – implementovaný prostřednictvím výpočetních operací konekcionistického systému, tedy lidského mozku. Podle implementacionalistů je tedy tou správnou úrovní

¹ „In a normal program, you can account for every single operation at the bit level, by looking “upward” toward the top-level program. You can trace a high-level function... calls subroutines that call other subroutines that call this particular machine-language routine... So there is a high-level, global reason why this particular bit is being manipulated.“

² Implementační detaily mohou hrát roli v případě, že by zásadním způsobem kolidovaly s časovými a jinými omezeními danými kontextem, ve kterém se daný výpočet realizuje.

abstrakce k výpočetní specifikaci mentálních procesů ta úroveň lidského kognitivního aparátu, která je dostupná vědomé introspekci a která je předmětem běžného mentalistického diskurzu. A stejně jako lze vytvořit výpočetní algoritmus pro digitální počítač, aniž bychom se přitom zabývali konkrétními detaily toho, jak tento počítač daný algoritmus vykoná (tj. prostřednictvím jakých základních operací ho provede), lze podle implementacionalistů také specifikovat výpočetní algoritmy mentálního výpočtu, aniž bychom se přitom museli starat o to, jak jsou tyto algoritmy konkrétně realizovány prostřednictvím primitivních operací biologického „wetwaru“ lidského mozku. Hofstadter (1985, s. 653) tento přístup charakterizuje jako postupující směrem odshora-dolů: *„funkce volající podfunkce volající podpodfunkce atd. až do chvíle, kdy se vše zarazí o dno tvořené nějakými primitivními operacemi. Intelligence takto byla [výzkumníky v oblasti umělé inteligence] považována za něco, co lze hierarchicky rozložit, s vyššími kognitivními funkcemi na vrcholu [hierarchie] řídícími nižší kognitivní funkce na dně [hierarchie].“*¹

Hofstadter (2000, s. 302-309) také nabízí zajímavou „meteorologickou“ metaforu, která výstižně zachycuje podstatu sporu mezi eliminativisty a implementacionalisty: Diskurz folkové meteorologie, který lidé běžně používají k popisu počasí, odkazuje k objektům jako jsou mraky, déšť, duha, vítr, oblačnost nebo sněhové vločky. Většina lidí si je přitom velice dobře vědoma, že jednotlivé mraky nebo déšť jsou pouze nápadným projevem různých dějů, které se týkají jiných objektů než jsou ty, o kterých se v diskurzu folkové meteorologie obvykle hovoří. Lidé tedy takto přijímají jako samozřejmost, že mezi folkovou meteorologií a fyzikou, resp. vědeckou meteorologií existuje určitá netriviální, ale přitom nijak magická propast, kterou dennodenně překračují oběma směry, když sledují televizní předpověď počasí; vědí, že vědecká meteorologie ve snaze objasnit a předpovídat různé meteorologické jevy na makroskopické úrovni vytváří různé výpočetní modely, které pracují s objekty, které jsou na nižší úrovni (abstrakce) než ty, na které lidé běžně odkazují v rámci laického meteorologického diskurzu; vědecká meteorologie se takto při modelování vývoje počasí opírá například o informace o teplotě, tlaku či vlhkosti vzduchu, nikoli o samotné objekty, které se snaží předpovídat. Tzn., že vědecká meteorologie při objasňování a předpovídání meteorologických jevů na makroskopické úrovni (tj. na úrovni srážek, oblačnosti nebo síly větru) nepracuje se zákonitostmi týkajícími se mraků, větrů a dalších objektů odpovídajících

¹ „functions calling subfunctions calling subsubfunctions and so on, until it all bottomed out in some primitives. Thus intelligence was thought to be hierarchically decomposable, with high-level cognition at the top driving low-level cognition at the bottom.“

běžným pojmům laické meteorologie, ale se zákonitostmi, které se týkají objektů na nižší, mikroskopické úrovni. To však neznamená, že by na makroskopické úrovni žádné zákonitosti neexistovaly. Zcela určitě existují a člověk (a nejen on) je také v mnoha různých situacích využívá. Podstatné je to, že můžeme uznat existenci zákonitostí na této makroskopické úrovni, a nepředpokládat přitom automaticky existenci nějakého formálního systému pravidel (definovaných na objektech makroskopické úrovně), kterými by se tyto zákonitosti řídily: Četným pozorováním makroskopických jevů jako jsou deště, mraky apod. lze objevit řadu zákonitostí a ty pak formulovat v podobě řady formálních pravidel popisujících jejich chování. Nic nám ale nezaručí, že na úrovni makroskopických objektů existuje taková sada formálních pravidel, která by nám umožnila přesně modelovat chování těchto makroskopických objektů. Možná že taková pravidla vůbec neexistují; možná že chování makroskopických meteorologických jevů prostě nelze plně vysvětlit a popsat jako výpočet (tedy jako soubor nějakých formálních pravidel) v pojmech jejich vlastní úrovně abstrakce. Poučení, které z toho podle Hofstadtera pro kognitivní vědu plyne, spočívá v tom, že ze skutečnosti, že úroveň mentalistického diskurzu je v mnoha případech vhodnou úrovní abstrakce k explanaci řady mentálních jevů, automaticky nevyplývá, že by na této úrovni (nebo na jiné jí blízké úrovni) měla být také formulována výpočetní teorie tyto mentální jevy vysvětlující: „Kdo může říci, zda na úrovni samotných aktivačních vzorců (čti „pojmu“, „idejí“, „myšlenek“) existují pravidla – formální, výpočetní pravidla –, která dokáží plně predikovat jejich tok... [a která] umožňují ignorovat vše, co se děje na nižší úrovni, a přitom stále poskytovat dokonalé nebo přinejmenším velice přesné předpovědi ohledně chování aktivačních vzorců. Jistě, existuje spousta fenomenologických pozorování, která mohou být formalizována tak, aby vypadala jako pravidla, která budou velice neurčitě popisovat chování těchto vysokoúrovňových aktivačních vzorců. Co nám ale zaručí, že se nám podaří získat veškerou flexibilitu vysokoúrovňové aktivity mozku a zapouzdřit ji – bez jakéhokoli nízkoúrovňového substrátu – do podoby nějakých výpočetních pravidel?“¹ (Hofstadter, 1985, s. 647) Podle implementacionalistů něco takového možné je. Hofstadter (1985, s. 654) tento postoj nazývá *booleovským snem* podle anglického logika George Boolea (viz s. 44-46), který byl přesvědčen, že zákony lidského myšlení spočívají ve formálních pravidlech pro

¹ „who can say whether there exist rules-formal, computational rules-at the level of the teams themselves (read “concepts”, “ideas”, “thoughts”) that are of full predictive power in describing how they will flow? I am speaking of rules that allow you to ignore what is going on “down below”, yet that still yield perfect or at least very accurate predictions of the teams’ behavior. To be sure, there are phenomenological observations that can be formalized to sound like rules that will describe, very vaguely, how those highest-level teams act. But what guarantee is there that we can skim off the full fluidity of the top-level activity of a brain and encapsulate it-without any lower substrate-in the form of some computational rules?“

manipulaci logických výroků (tj. výroků, které jsou buď pravdivé, nebo nepravdivé), tedy že proud myšlenek (aktivačních vzorců) se řídí pravidly, která jsou definována na úrovni samotných myšlenek (aktivačních vzorců). Eliminativisté jsou naproti tomu přesvědčeni, že výpočetní teorie vysvětlující chování aktivačních vzorců (myšlenek), které jsou nositeli intencionality (či sémantiky), se kterou pracuje mentalistický diskurz, musí být formulovány na nižší úrovni než je ta, která odpovídá běžným myšlenkám a pojmům. Myšlení je tak podle nich určitým statisticko-emergentním jevem ve smyslu „nenaprogramované funkčnosti“, která nepřímo povstává z paralelní aktivity velkého množství subsymbolových prvků konekcionistického systému. V této souvislosti je užitečný filozofický koncept rozdílu mezi systémy, které se řídí určitými pravidly (*rule following behavior*), a systémy, které se samy o sobě žádnými pravidly neřídí, ale jejichž chování lze přesto prostřednictvím určitých pravidel relativně přesně popsat (*rule described behavior*). Často uváděným příkladem jsou planety, jejichž pohyb lze s pomocí určitých pravidel velice přesně popsat a vypočítat, a to přesto, že se planety samy o sobě žádnými pravidly neřídí a svoje oběžné dráhy nijak nepočítají. Toto pravidly popsatelné chování planet je důsledkem chování jiných objektů (atomů, molekul), které se již určitými pravidly skutečně řídí. Pravidelnosti v chování planet jsou takto pouhým epifenomémem pravidel, kterými se řídí objekty na nižší úrovni analýzy než jsou samotné planety. Stejnou úvahu pak lze aplikovat i na vztah symbolových a konekcionistických systémů: Tok aktivačních vzorců (myšlenek tvořících obsah vědomého a introspekci nahlédnutelného myšlení) lze často popsat prostřednictvím souboru pravidel, kterými se však aktivační vzorce samy o sobě neřídí; pravidelnosti v jejich toku jsou pouze epifenomémem pravidel, kterými se na nižší úrovni analýzy skutečně řídí (tj. jakoby „konzultují“ s nimi svoji činnost) jednotlivé funkční jednotky konekcionistického systému. Tento rozdíl lze v ideálních podmínkách často „beztrestně“ ignorovat; existují ovšem situace, kdy se výběr špatné úrovně abstrakce k formulování výpočetní (formální) teorie daného jevu zřetelně projeví.

Tento spor mezi eliminativisty a implementacionalisty lze bez nějakého velkého interpretačního násilí přirovnat ke středověkému filozofickému sporu o univerzálie. V obou dvou případech se jedná o spor ohledně povahy určité třídy pojmů; rozdíl spočívá pouze v tom, o jakou třídu pojmů se jedná a o jakou jejich povahu se spor vede: Ve středověku se spor vedl mezi *realisty* a *nominalisty* o to, zda obecné pojmy (univerzálie) jako „pes“ či „židle“ mají svojí vlastní nezávislou existenci, tedy zda skutečně existují, podobně jako platónské ideje (realistická pozice), nebo zda se jedná pouze o abstrakta či jména, která existují jen v lidských myslích jako užitečné nálepky, které ekonomizují potýkání se s nekonečnou rozmanitostí světa předmětů a událostí (nominalistická pozice) (Blecha, 1996).

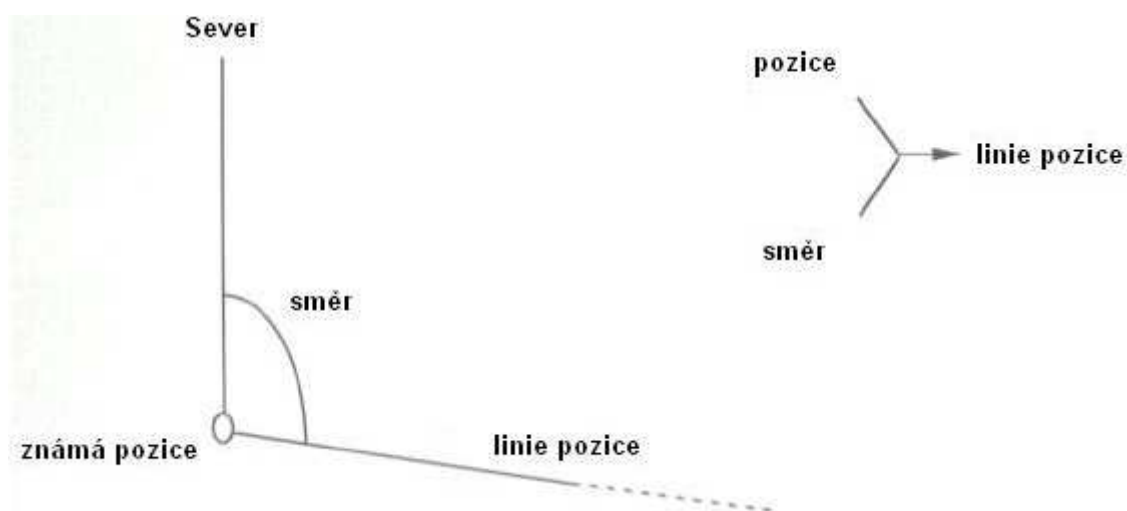
Ve sporu eliminativistů a implementacionalistů pak jde o rozdíl v názoru na to, zda pojmy mentalistického diskurzu mají či nemají výpočetní povahu, tedy zda vystupují jako svébytné objekty operací mentálního výpočtu, nebo jestli se jedná pouze o více či méně užitečné nálepky, které jsou jen přibližným popisem toho, co se děje během mentálního výpočtu.

Vedle tradičního eliminativistického a implementacionalistického přístupu zde existuje ještě „třetí cesta“, která vychází z předpokladu, že lidská mysl má hybridní povahu a že se tedy skládá zároveň jak z čistě konekcionistického systému tak i ze systému čistě symbolového. Příkladem takového přístupu je Smolenskyho (1988) koncepce dvou druhů mentálních procesorů. Podle Smolenskyho má primárně velká část poznatků, se kterými člověk pracuje, povahu lingvistických reprezentací, které Smolensky nazývá „kulturními poznatky“ (*cultural knowledge*). Tyto poznatky pak lidé využívají prostřednictvím určitého virtuálního stroje (implementovaného na konekcionistickém hardwaru), který manipuluje vnitřními korelátory těchto lingvistických reprezentací v souladu s pravidly definujícími danou problémovou doménu. Tento virtuální stroj Smolensky ztotožňuje se sériovým, pomalým a vědomým myšlením, které všichni dobře znají na základě introspekce, a nazývá ho *vědomým interpretem pravidel* (*conscious rule interpreter*). Druhým typem mentálního procesoru je tzv. *intuitivní procesor* (*intuitive processor*). Rozdíl mezi oběma procesory spočívá v objektech, se kterými pracují. Zatímco syntaktickými objekty výpočetních operací vědomého procesoru jsou lingvistické reprezentace používané k popisu různých problémových domén, syntaktickými objekty intuitivního procesoru jsou distribuované reprezentace různých „mikrocharakteristik“, které mají k sémantickým entitám běžného diskurzu vztah poněkud volnější. To jinými slovy znamená, že vědomý procesor je na rozdíl od intuitivního procesoru sémanticky transparentní: syntax a sémantika programů, které běží na vědomém procesoru, těsně korespondují s lingvistickým popisem problémových domén, což v případě programů běžících na intuitivním procesoru neplatí. Vědomý procesor je podle Smolenskyho odpovědný za specificky lidské kognitivní schopnosti spadající do oblasti abstraktního myšlení, zatímco intuitivní procesor, který je odpovědný za „*percepci, motoriku, plynulé jazykové chování, intuici při řešení problémů a hraní her... zkrátka prakticky za veškerý zkušený a rutinní výkon*“¹ (Smolensky, 1988, s. 5), má člověk společný z řadou dalších, vývojově nižších živočichů.

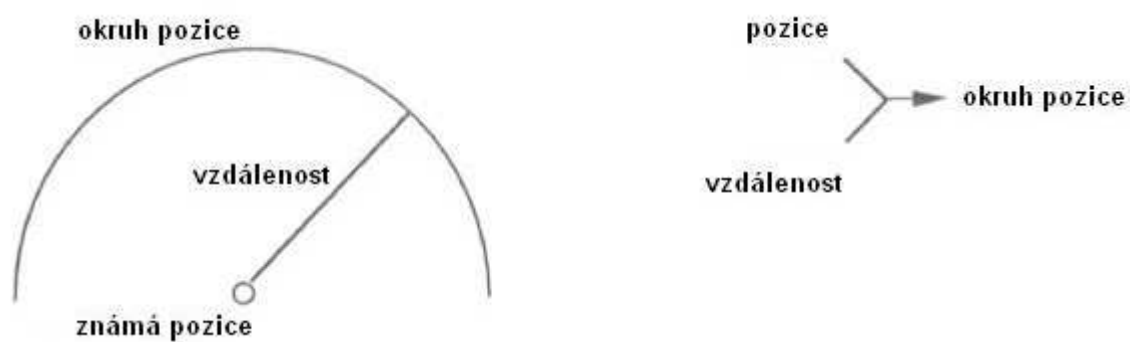
Přechod od sémanticky transparentního zpracovávání informací (vědomého procesoru) k sémanticky netransparentnímu zpracovávání informací (intuitivního procesoru) přitom

¹ „...perception, practiced motor behavior, fluent linguistic behavior, intuition in problem solving and game-playing. In short, practically all of skilled performance.“

podle Smolenskyho není náhlý (ve stylu buď/anebo), ale spíše pozvolný, kdy subsymbolový (intuitivní) procesor (který tvoří základní výpočetní architekturu lidského kognitivního systému) může v různých situacích více či méně úspěšně aproximovat chování symbolového (vědomého) procesoru, který manipuluje syntaktickými objekty odpovídajícími pojmům běžného diskurzu. Čím větší je dimenzionální posun (viz s. 94-95) mezi sémantikou prvků konekcionistického systému a sémantikou objektů symbolového systému, tím je tato aproximace problematičtější a jen velice přibližná. Podle Smolenskyho lze tento koncept (aproximace chování symbolového systému systémem konekcionistickým) uchopit nejlépe tak, že budeme tradiční symbolové modely kognitivních funkcí považovat za pouhý popis kompetence kognitivního systému, tedy za popis schopnosti kognitivního systému řešit určité problémy za ideálních podmínek, kdy má kognitivní systém k dispozici všechny potřebné vstupní informace a kdy má na jejich zpracování neomezené množství času. Tento popis kompetence kognitivního systému má podobu specifikace tzv. „tvrdých“ omezení (*hard constraints*) výpočetního procesu v pojmech běžného diskurzu. Smolensky jako příklad takových tvrdých omezení uvádí jednoduchou rovnici Ohmova zákona U (*elektrické napětí*) $= I$ (*elektrický proud*) $\times R$ (*elektrický odpor*), kterou se v nějaké podobě musí řídit každý kognitivní systém, jehož úkolem je dávat odpovědi na otázky typu *Jestliže se zvýší elektrický odpor, jaký vliv to bude mít na hodnotu elektrického napětí? Zvýší se, sníží se, nebo zůstane stejná?* Jiným příkladem by mohla být omezení, kterým musí vyhovět jakýkoli navigační systém, jehož základní funkcí je odpovídat na otázku typu *Kde se právě nacházím?* Hutchins (1995) takto hovoří o dvou základních omezeních, jejichž kombinací (současným respektováním) navigační systém plní svou základní funkci: Prvním omezením je tzv. linie pozice (*line of position*), která vzniká kombinací známé pozice a daného směru; takže jestliže vím, že bod B se nachází v určitém směru od známé pozice bodu A, potom vím, že bod B se bude nacházet někde na přímce vedené v daném směru z bodu A (viz obrázek 39). Samotná znalost linie pozice ovšem k přesné lokalizaci nestačí. Je zapotřebí dalšího omezení, kterým je tzv. okruh pozice (*circle of position*), který vzniká kombinací známé pozice a udané vzdálenosti; jestliže vím, že bod B se nachází v určité vzdálenosti od známé pozice bodu A, potom vím, že bod B se bude nacházet někde na obvodu kružnice s daným poloměrem a středem v bodu A (viz obrázek 40). Různou kombinací těchto dvou jednodimenzionálních omezení lze vytvořit dvojdimenzionální omezení, které přesně lokalizuje pozici na 2D ploše (viz obrázek 41). Navigační systém kromě otázky, kde se člověk právě nachází, odpovídá rovněž na otázky typu *Jsme-li právě tam, kde jsme, jak dál pokračovat, abychom se dostali na požadovanou pozici?* nebo *Jsme-li jsme právě tam, kde*



Obrázek 39: Grafické a schematické znázornění omezení linie pozice (line of position constraint). (s úpravami převzato z Hutchins, 1995, s. 52)

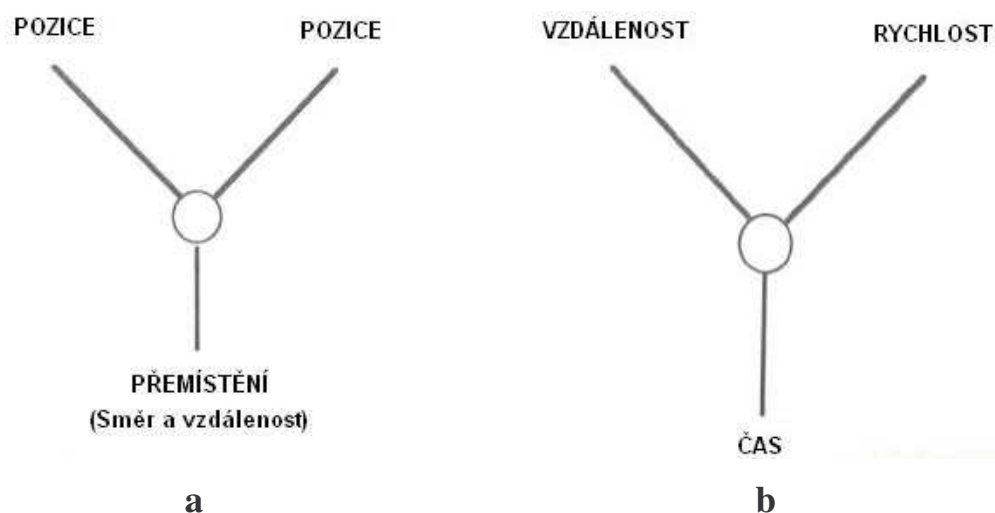


Obrázek 40: Grafické a schematické znázornění omezení okruhu pozice (circle of position constraint). (s úpravami převzato z Hutchins, 1995, s. 53)



Obrázek 40: Schematické znázornění různých způsobů kombinace jednodimenzionálních omezení. (s úpravami převzato z Hutchins, 1995, s. 53)

jsme, kam se dostaneme, jestliže budeme postupovat daným způsobem? U první otázky se používají informace o lokalizaci dvou míst k určení pozičního vztahu (směru a vzdálenosti) mezi nimi; u druhé otázky se pak používají informace o lokalizaci jednoho místa a o pozičním vztahu k určení lokalizace druhého místa. V případě obou otázek se tedy jedná o poziční vztah mezi dvěma místy (tzv. omezení pozice a přesunutí, *position-displacement constraint*). Vztah mezi těmito třemi omezeními lze vyjádřit pomocí jednoduchého schématu (viz obrázek 40a), kde specifikace jakýchkoli dvou položek umožňuje přesné určení i třetí položky. Poslední omezení, se kterým navigační systém pracuje, se týká vztahu mezi časem, rychlostí a vzdáleností (*distance-time-rate constraint*, viz obrázek 40b). Stejně jako v případě omezení pozice a přesunutí i zde znalost jakýchkoli dvou položek jednoznačně specifikuje hodnotu třetí položky. Toto omezení navigačnímu systému umožňuje odpovídat na otázky typu *Jestliže za 1 hodinu urazím v průměru 60 km, jak daleko se dostanu za 6 hodin nepřetržité plavby?* nebo *Jak dlouho mi bude trvat cesta do 100 km vzdáleného města, jestliže ujedu 25 km za hodinu?*



Obrázek 41: a) Schematické znázornění omezení pozice a přesunutí (*position-displacement constraint*). b) Schematické znázornění omezení rychlosti, vzdálenosti a času (*distance-rate-time constraint*). (s úpravami převzato z Hutchins, 1995, s. 55, 57)

Za ideálních podmínek, kdy jsou k dispozici všechny potřebné vstupní informace a kdy je na jejich zpracování neomezené množství času, se bude chování symbolového a konekcionistického systému shodovat. Za jiných než ideálních podmínek se však chování obou systémů bude lišit, neboť přestože je v ideálním případě chování konekcionistického systému popsatelné jako vyhovující tvrdým omezením, ve skutečnosti konekcionistický systém své výpočty provádí tak, že respektuje velké množství tzv. „měkkých“ omezení (*soft*

constraints) v podobě numericky kódovaných vah spojů mezi uzly sítě, jejichž sémantika přímočaře nekorresponduje s pojmy, kterými je obvykle charakterizována daná problémová doména, resp. kompetence řešit problémy v této oblasti. Tento rozdíl se může projevit například tak, že zatímco symbolový systém při absenci či poškození některých vstupních informací pozbude své kompetence řešit daný typ problémů, konekcionistický systém si i v těchto neideálních podmínkách uchová schopnost vyprodukovat nějaký „rozumný“ výstup (viz také odolnost konekcionistických systémů vůči chybným nebo neúplným vstupním informacím na stranách 91 a 93). Smolensky tento vztah přirovnává ke vztahu mezi klasickou, newtonovskou fyzikou a kvantovou fyzikou: Fyzikální svět je ve své podstatě kvantový systém, který se za určitých podmínek jeví být newtonovským systémem (kde se objekty nacházejí vždy právě na jednom místě...); podobně je kognitivní systém ve své podstatě konekcionistickým systémem, který se za určitých podmínek (nad hranicí 100 milisekund?) jeví být systémem symbolovým. Podobně Rumelhart a McClelland (1986c, s. 125) tvrdí, že *„konvenční symbolové procesní systémy jsou makroskopickými popisy, které jsou analogiemi newtonovské mechaniky, zatímco...[konekcionistické] modely nabízejí více mikroskopický popis, který je analogií kvantové teorie... [a] makroskopická úroveň popisu může být pouze aproximací teorie, která je více mikroskopická.“*¹ *„Rozdíl mezi oběma popisy je [přitom] patrný až v extrémních podmínkách..., takže v mnoha případech je klasický [symbolický] pohled dobře fungující aproximací. Jenže funkčnost a skutečnost jsou dvě odlišné věci.“* (Greene, 2006, s. 20)

1.2.3.5 Distribuované kognitivní systémy

Poněkud odlišný pohled na povahu výpočetní architektury lidské mysli nabízí **teorie distribuované kognice** (*theory of distributed cognition*), podle které se funkční prvky lidského kognitivního systému nenacházejí pouze uvnitř mozku, ale také ve vnějším prostředí, například v podobě různých kognitivních artefaktů (viz také oddíl „1.2.1.2 Historický a kognitivně antropologický kontext formálních systémů“). Právě tato distribuce výpočetních procesů mezi biologické a nebiologické struktury má podle této teorie vysvětlovat vysoce specifický kognitivní profil člověka. Ten na jedné straně exceluje v takových úlohách jako je senzomotorická koordinace, rozpoznávání různých objektů, orientace v prostoru apod. a na

¹ „...conventional symbol processing models are macroscopic accounts, analogous to Newtonian mechanics, whereas our models offer more microscopic accounts, analogous to quantum theory... Through a thorough understanding of the relationship between the Newtonian mechanics and quantum theory we can understand that the macroscopic level of description may be only an approximation to the more microscopic theory.“

straně druhé je schopen (ovšem již s jistým sebezapřením a s vynaložením nemalé duševní námahy a úsilí) abstraktního a sekvenčního myšlení, například v podobě plánování posloupnosti dílčích kroků vedoucích k nějakému vzdálenému cíli. Jak již bylo uvedeno v předchozích dvou oddílech, k modelování vědomého sériového usuzování jsou vhodné především klasické symbolové systémy, zatímco k modelování (většinou) nevědomých kognitivních procesů spojených se senzomotorickou koordinací, s percepcí a rozpoznáváním různých objektů jsou vhodnější konekcionistické systémy. Vzhledem k tomu, že biologický substrát lidského kognitivního systému je podle všeho konekcionistickým systémem, vzniká tak problém, jak z konekcionistického stroje „dostat“ něco jako je vědomé sériové usuzování, které člověk běžně používá při řešení různých abstraktních problémů.

Podle teorie distribuované kognice řešení tohoto problému spočívá v **propojení dynamiky konekcionistického systému s dynamikou a strukturou vnějšího prostředí**. Toto propojení lze ilustrovat na příkladě procedury dlouhého násobení s pomocí tužky, papíru a symbolické notace: Většina lidí se bez větších problémů dokáže naučit nazpaměť výsledky malé násobilky, takže v případě zadání úlohy jako je „6 x 7“ dokáží velice rychle a bez nějakého velkého přemýšlení přijít se správnou odpovědí („42“). Tuto schopnost lze snadno realizovat prostřednictvím konekcionistického systému, který je schopen velice úspěšně rozpoznávat a kompletovat různé neúplné vzorce a struktury nacházející se ve vnějším prostředí. Násobení větších čísel, např. „7654 x 563“, však již představuje mnohem obtížnější problém, při jehož řešení si lidé obvykle pomáhají tím, že používají tužku, papír a nějakou symbolickou notaci k zápisu dílčích mezivýsledků malé násobilky („3 x 4“, „3 x 5“...), jejichž součtem se člověk dobere konečného řešení. Lidé takto komplexní problém zredukuje na posloupnost několika jednodušších problémů, při jejichž řešení si vystačí se schopnostmi, které podporuje jejich základní biologický hardware. To, co většině lidí umožňuje si poradit s násobením velkých čísel, je tak tedy propojení jejich základního biologického hardwaru s externím paměťovým médiem. Rumelhart a kol. (1986a, s. 46) k tomu poznamenávají, že *„toto je skutečné zpracovávání symbolů..., primární zpracovávání symbolů, kterého jsme schopni. Z tohoto pohledu se vnější prostředí stává klíčovým rozšířením naší mysli.“*¹

Častým počítáním podobných příkladů se člověk může časem naučit provádět takové výpočty i tzv. „z hlavy“, tedy již bez pomoci tužky a papíru. Tuto schopnost počítat „z hlavy“ umožňuje mentální simulace té činnosti, kterou člověk dříve prováděl s pomocí vlastních očí a rukou: Opakovaným prováděním výpočtů na papíře si člověk často nakonec osvojí

¹ „This is real symbol processing and, we are beginning to think, the primary symbol processing that we are able to do. Indeed, on this view the external environment becomes a key extension to our mind.“

schopnost manipulovat ve své představivosti mentálním modelem symbolického zápisu mezivýsledků stejným způsobem, jakým dříve manipuloval symboly skutečně zapsanými na kusu papíru. Touto mentální simulací procesů ve vnějším prostředí tak člověk dokáže internalizovat kognitivní funkce, které mají jinak svůj zdroj v manipulaci s reálnými objekty ve vnějším prostředí. Důležitá je zde ta skutečnost, že tato vnitřní mentální manipulace se symboly má jinou povahu než manipulace s vnitřními symboly, o které se hovoří v souvislosti s tradičními symbolovými systémy: Zatímco v případě symbolových systémů je manipulace s vnitřními symboly považována za samotnou podstatu výpočetní architektury lidské mysli, v případě mentálně simulované manipulace s vnějšími symboly se jedná pouze o simulaci jistých procesů ve vnějším prostředí, která je realizována konekcionistickým výpočetním systémem (tj. lidským mozkem).

Podle zastánců teorie distribuované kognice je celá tradiční kognitivní věda založena na chybné a zavádějící projekci kognitivních vlastností celého distribuovaného kognitivního systému - sestávajícího z biologického výpočetního hardwaru a ze struktury a dynamiky vnějšího prostředí - pouze do jedné jeho komponenty, totiž do biologického výpočetního hardwaru. Hutchins (1995) upozorňuje, že zdroj této chyby lze vysledovat až k Alanu Turingovi, konkrétně k jeho práci na formalizaci pojmu výpočtu v podobě Turingova stroje, jehož výpočetní architektura (konečný automat + nekonečně dlouhá paměťová páska) se stala předobrazem výpočetní architektury moderního digitálního počítače (CPU + paměť). Ten v tradiční kognitivní vědě hraje roli paradigmatického modelu výpočetních operací lidské mysli. To, že počítač může tuto roli hrát, je podle představitelů tradiční kognitivní vědy dáno tím, že „*počítač byl vytvořen k obrazu člověku*“¹ (Simon, Kaplan, 1989). Pomyslným demiurgem, který je za tento „akt stvoření“ zodpovědný, je právě Turing, který si na základě pečlivé introspektivní analýzy posloupnosti svých vlastních myšlenkových pochodů během provádění numerických výpočtů a řešení různých matematických problémů uvědomil, že výpočet – jakkoli komplikovaný - je pouze postupnou a systematickou aplikací několika málo primitivních manipulačních operací na symbolické struktury. Tento svůj náhled Turing vtělil do imaginárního výpočetního zařízení (Turingova stroje), které dokáže všechny tyto operace provádět zcela automaticky bez jakéhokoli zásahu lidského operátora. Funkční architektura tohoto imaginárního zařízení se později stala přímou inspirací pro konstrukci moderního digitálního počítače. Díky tomuto „rodokmenu“ se počítač může pyšnit hrdým titulem „*výsadního modelu lidské mysli*“. Problém s tímto zakladatelským mýtem spočívá v tom, že

¹ „*The computer was made in the image of human.*“

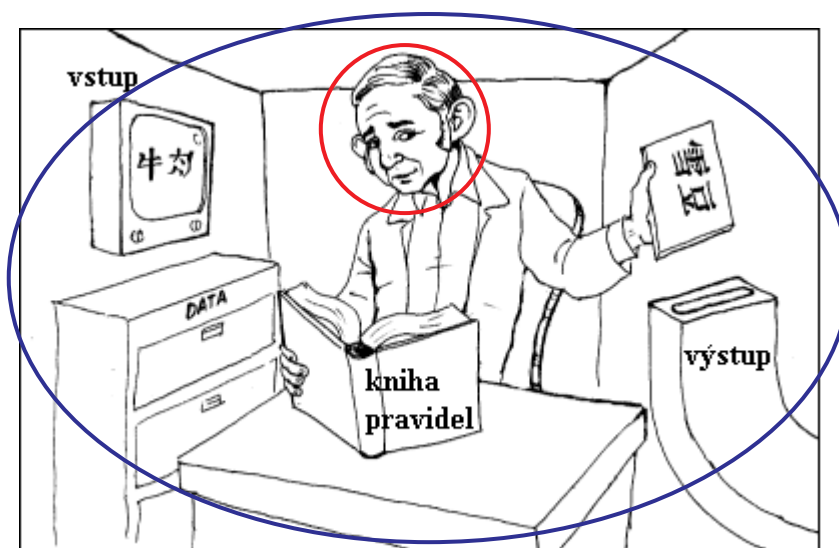
skutečným kognitivním systémem, který stál počítači modelem, byla fyzická osoba, která fyzicky manipulovala symboly na kusu papíru s pomocí svých rukou a očí. A právě tato vizuální a manuální interakce matematika či logika s hmotným světem (v podobě různých symbolických struktur) prováděla daný výpočet, nikoli samotný matematik nebo logik. Jinými slovy, máme-li zde nějaké symboly a člověka, který s nimi fyzicky manipuluje, potom jsou kognitivní operace, které člověk během fyzické manipulace symbolů provádí, naprosto odlišné od kognitivních operací celého systému, který se skládá z člověka a symbolických struktur, se kterými člověk ve vnějším prostředí manipuluje: „*Vlastnosti člověka v interakci se symboly realizují určitý druh výpočtu. To ale neznamená, že by se tento výpočet odehrával také uvnitř hlavy tohoto člověka.*“¹ (Hutchins, 1995, s. 361) Hutchins (1995, s. 361-362) jako ilustraci tohoto efektu uvádí Searlův „Čínský pokoj“ (viz obrázek 42) - myšlenkový experiment, který Searle původně vytvořil jako argument pro podporu svého tvrzení, že dodržování syntaktických pravidel nepostačuje k reprodukci všech sémantických charakteristik mentálního života člověka. Podstata tohoto myšlenkového experimentu je následující: John Searle, který umí pouze anglicky, je uzavřen v nějaké místnosti. Skrze tenký otvor ve dveřích mohou čínští rodilí mluvčí do této místnosti odesílat různé řetězce čínských znaků, které mohou znamenat jakékoli tvrzení či otázku v čínském jazyce. Stejným otvorem může Searle zase posílat řetězce čínských znaků ven. Searle přitom dokáže na dotazy a



Obrázek 42: Searlův „Čínský pokoj“.

¹ „The properties of the human in interaction with the symbols produce some kind of computation. But that does not mean that computation is happening inside the person's head.“

tvrzení čínských rodilých mluvčích reagovat sestavením a zasláním takových řetězců čínských znaků, které vedou čínské rodilé mluvčí k přesvědčení, že v místnosti se musí nacházet někdo, kdo čínštině opravdu dobře rozumí. Searle však čínštině ve skutečnosti vůbec nerozumí; čínské znaky jsou pro něj pouze nesmyslnými klikyháky. Avšak Searle má v místnosti vedle tří košů s čínskými znaky také knihu pravidel, která mu dává jednoznačné instrukce, jaké řetězce čínských znaků má vybrat a jakým způsobem je má sestavit a následně odeslat v reakci na přijetí jiných řetězců čínských znaků z vnějšího světa. Podle Searla tento myšlenkový experiment¹ dokazuje, že přestože se pokoj jako celek chová způsobem, který budí dojem, že rozumí čínštině, ve skutečnosti pokoj, ani nic jiného v něm čínštině vůbec nerozumí. Hutchins však tento myšlenkový experiment využívá k poněkud odlišnému účelu než Searle. Podle něj lze na něm velice dobře ilustrovat to, jak se kognitivní charakteristiky distribuovaného kognitivního systému mohou lišit od kognitivních vlastností jedné jeho komponenty: Zatímco samotný John Searle neumí čínsky ani slovo, pokoj jako celek, tj. Searle + čínské znaky + kniha pravidel, se zdá čínštině rozumět.



Obrázek 43: Searlův „Čínský pokoj“ názorně ilustruje, jak zásadním způsobem se mohou lišit kognitivní charakteristiky distribuovaného kognitivního systému jako celku a kognitivní charakteristiky jedné jeho dílčí komponenty.

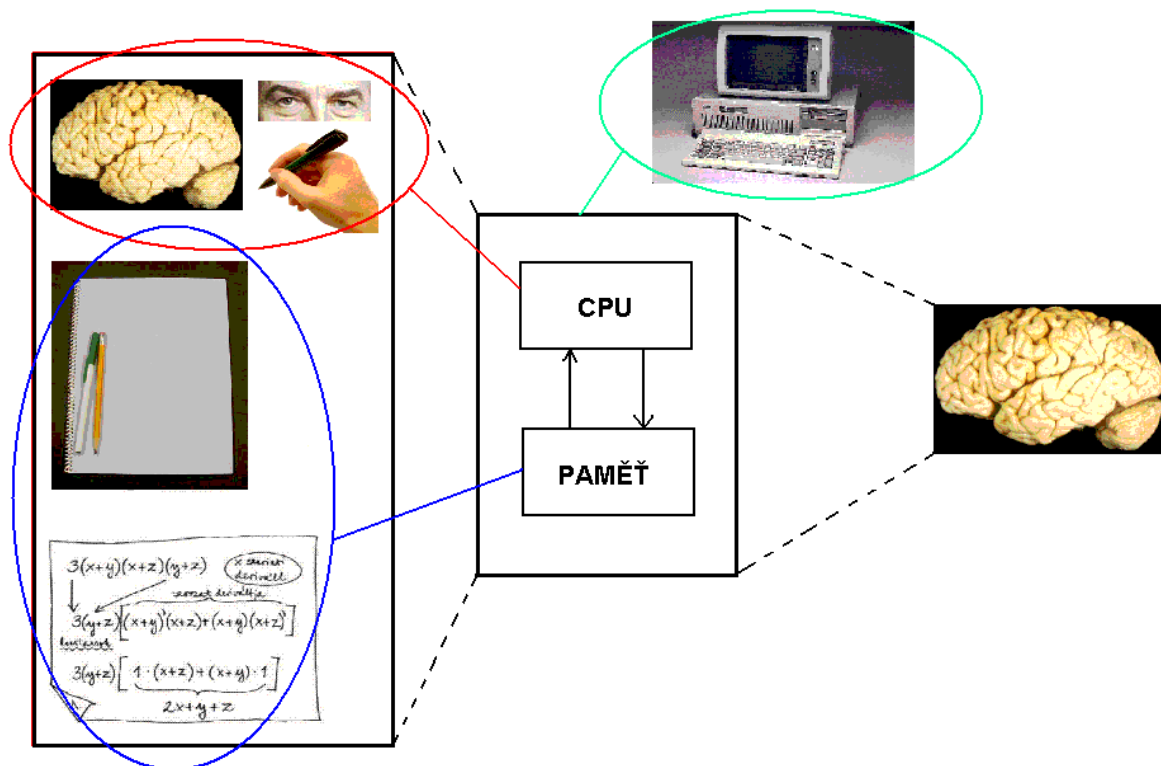
¹ Jakousi dobu Searlova myšlenkového experimentu skutečně zrealizovali Newell se Simonem, když svůj první počítačový program modelující lidskou kognici, *Logického teoretika* (*Logic Theorist*), implementovali prostřednictvím skupiny lidí, z nichž každý měl na starost realizaci nějakého jednoduchého podprogramu zapsaného na malé kartičce: „Každému členovi skupiny jsme dali jednu z kartiček, takže každá osoba se stala součástí počítačového programu Logického teoretika – podprogramem, který plnil určitou speciální funkci, nebo byl prvkem jeho paměti. Úkolem každé účastníka bylo vykonat svůj podprogram nebo dát k dispozici obsah své paměti, kdykoli byl volán procedurou o úroveň výš, která byla zrovna aktivní. Takto jsme byli schopni simulovat chování Logického teoretika s pomocí počítače sestaveného z lidských součástek... Jednotlivé osoby nebyly o nic více zodpovědné... než Platónův Menon, přesto byly [tyto osoby] úspěšné při dokazování teorémů, které jim byly předloženy.“ (Simon, 1991, s. 207)

Nazíráno z této perspektivy se pak zdá, že podstata Turingova objevu spočívá v tom, že je možné abstrahovat od konkrétního způsobu, jakým skutečný matematik realizuje výpočet (totiž jako fyzickou manipulaci symbolů prostřednictvím svých očí a rukou), a že to podstatné je pouze systematická aplikace transformačních operací na řetězce symbolů. To potom znamená, že matematika lze z celého výpočetního procesu odstranit a celý proces manipulace symbolů zautomatizovat prostřednictvím nějakého výpočetního stroje (počítače). Ten ovšem nemodeluje ty kognitivní procesy, které matematik skutečně používá při fyzické manipulaci symbolů nacházejících se ve vnějším prostředí, ale modeluje pouze abstraktní manipulaci symbolů, resp. to, co vykonává distribuovaný kognitivní systém jako celek, který se skládá z matematika a nějakého formálního systému, se kterým matematik podle určitých pravidel manipuluje (a tak něco počítá). Tzn. že počítač není modelem kognitivních procesů individuální lidské mysli, ale širšího, distribuovaného kognitivního systému: „*Matematik,...který byl v interakci s hmotným světem, není [počítačový] systémem ani modelován, ani v něm není ničím jiným nahrazen. Tato osoba v systému automatické manipulace symbolů jednoduše chybí. To, co je modelováno, je abstraktní výpočet realizovaný manipulací symbolů. [...] Architektura fyzikálního symbolového systému [tak] není modelem individuální kognice [ale] modelem funkce sociokulturního systému, ze kterého byl odstraněn lidský činitel.*“¹ (Hutchins, 1995, s. 363) To, že si tradiční kognitivní věda tohoto nebyla vědoma a že považovala počítač za přiměřený model lidské mysli, vedlo nakonec k tomu, že to, co dříve bylo vně kognitivního systému, tj. různé symbolické struktury, se najednou ocitlo uvnitř kognitivního systému a stalo samotným jádrem jeho výpočetní architektury (viz obrázek 44). Mozek, který prostřednictvím očí a rukou prováděl veškerou fyzickou manipulaci se symbolickými strukturami ve vnějším prostředí, tak byl nahrazen počítačem provádějícím manipulaci s „odtělesněnými“ symbolickými strukturami nacházejícími se uvnitř samotného kognitivního systému. Tímto došlo k tomu, že se symboly přesunuly z vnějšího prostředí dovnitř lidské mysli (nyní symbolového systému) a že se lidská mysl svým způsobem oddělila od vnějšího prostředí (nyní pouze zdroje vstupních informací specifikujících problémovou situaci a místa pro realizaci vypočítaného řešení problému): „*Symbolové systémy jsou vnitřním prostředím - chráněným před vnějším světem -, ve kterém*

¹ „*The mathematician who was a person interacting with a material world is neither modeled by this system nor replaced in it by something else. The person is simply absent from the system that performs automatic symbol manipulation. What is modeled is the abstract computation achieved by the manipulation of the symbols. [...] The physical-symbol system architecture is not a model of individual cognition. It is a model of the operation of a sociocultural system from which the human actor has been removed.*“

může probíhat zpracovávání informací ve službě [potřebám] organismu.“¹ (Newell a kol., 1989, s. 107)

To, co podle zastánců teorie distribuované kognice (Lave, 1988; Donald, 1991; Kirsh, Maglio, 1994; Hutchins, 1995; Clark, 1997) dělá lidský kognitivní profil jedinečným, je jeho schopnost vytvářet a využívat různé podpůrné struktury ve vnějším prostředí k transformaci povahy různých problémových prostorů, tak aby se tyto staly zvládnutelné i s pomocí relativně omezeného kognitivního profilu lidského mozku. Tyto podpůrné struktury se nazývají **kognitivní artefakty** či **nástroje mysli** (viz oddíl „1.2.1.2 Historický a kognitivně



Obrázek 44: Základní chyba tradiční kognitivní vědy, která předpokládá, že lidská mysl je druh symbolového procesoru, spočívá v tom, že zaměnila kognitivní profil distribuovaného kognitivního systému - skládajícího se z mozku a z řady podpůrných struktur vnějšího prostředí - za kognitivní profil samotného mozku. Výpočetní architektura moderního digitálního počítače, pro kterou je charakteristické striktní oddělení procesoru a paměti, odráží oddělení mozku od podpůrných struktur ve vnějším prostředí v distribuovaném kognitivním systému, jehož je počítač skutečným modelem. „Počítač [fakto] nebyl vytvořen k obrazu člověka...[, ale] k obrazu formálních manipulací abstraktních symbolů. A posledních 30 let kognitivní vědy lze chápat jako snahu předělat člověka k obrazu počítače.“² (Hutchins, 1995, s. 363)

¹ „Symbol systems are an interior milieu, protected from the external world, in which information processing in the service of the organism can proceed.“

² „The computer was not made in the image of the person. The computer was made in the image of the formal manipulations of abstract symbols. And the last 30 years of cognitive science can be seen as attempts to remake the person in the image of the computer.“

antropologický kontext formálních systémů“), které v kombinaci se základními výpočetními operacemi lidského mozku vytvářejí širší systém, který je schopen se úspěšně vypořádávat s problémy, které by jinak byly pro „holý“ mozek nezvladatelné; asi tak jako když člověk vybavený nůžkami najednou začne být schopen řešit úlohy, které doposud byly zcela mimo jeho dosah (například čistě a přesně rozpůlit list papíru). Jak k tomu poznamenává Dennett (2004b, s. 96), *„užívání nástrojů je dvousměrným znakem inteligence; nejenom že rozpoznat a ovládnout nástroj (o vyrobení ani nemluvě) vyžaduje inteligenci, ale nástroj i uděluje inteligenci těm, kdo mají to štěstí, že jim byl nějaký dán“*. Asi nejvýznamnějším nástrojem myslí je přitom jazyk; ten dává abstraktním vztahům a objektům podobu snadno identifikovatelných a manipulovatelných objektů, se kterými si dokáží dobře poradit vrozené schopnosti lidského mozku sledovat a rozpoznávat různé struktury, čímž jazyk zpřístupňuje jinak velice obtížně uchopitelné problémové domény (včetně vlastního myšlení, tj. myšlení o myšlení; viz také s. 35-36). Clark (1998a, s. 162) takto jazyk charakterizuje jako druh *„vnějšího artefaktu, jehož... adaptivní hodnota částečně spočívá v jeho roli, kterou hraje v transformaci těch druhů výpočetních prostorů, se kterými se naše biologické mozky musí potýkat při řešení určitých druhů problémů nebo při realizaci určitých komplexních projektů.“*¹

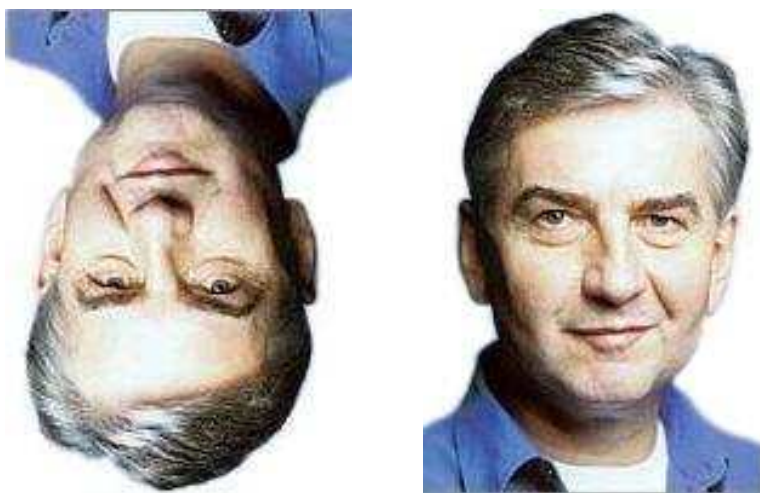
Toto pojetí řešení různých problémů jako procesu, ve kterém jsou úzce provázány vnitřní kognitivní operace s nejrůznějšími podpůrnými strukturami ve vnějším prostředí, je v ostrém rozporu s tradičním, odtělesněným pojetím plánování jako úplné vnitřní (mentální) specifikace posloupnosti činností, kterou stačí pouze zrealizovat ve vnějším světě, aby se dosáhlo vytčeného cíle. Plán řešení nějakého problému je tak v tradičním pojetí jakousi vnitřní obdobou sady instrukcí pro uvaření polévky či sestavení nějakého kusu nábytku a vnější prostředí zde slouží pouze jako zdroj vstupních informací, které specifikují daný problém, a jako prostor, kde se realizuje dopředu vytvořený plán posloupnosti činností vedoucích k danému cíli. Když se ale podíváme na způsob, jakým lidé skutečně plánují jednotlivé kroky při řešení různých problémů, uvidíme, že se zde téměř vždy vyskytuje komplexní souhra mezi vnitřním plánem činností a podpůrnými strukturami ve vnějším prostředí.

Tento rozdíl mezi dvěma přístupy k řešení problémů, k plánování a k lidské kognici obecně zachytili kognitivní vědci David Kirsh a Paul Maglio (1994) v konceptu tzv.

¹ „Public language... is a species of external artifact whose current adaptive value is partially constituted by its role in re-shaping the kinds of computational space that our biological brains must negotiate in order to solve certain types of problems, or to carry out certain complex projects.“

pragmatických a epistemických činností. Zatímco pragmatické činnosti jsou prováděny za tím účelem, aby se svět změnil co možná nejvíce směrem k žádoucímu stavu, tedy směrem k nějakému předem stanovenému cíli (například člověk si takto nejdříve musí sundat kabát, než si může sundat tričko), epistemické činnosti jsou činnostmi, jejich primárním účelem je změnit povahu úkolů, kterým lidská mysl čelí, tak aby pro ni byly co možná nejjednodušší. V obou dvou případech tedy člověk nějak fyzicky jedná ve světě, ale v případě epistemických činností jsou tou hlavní motivací kognitivní potřeby související se získáváním a zpracováváním informací.

O epistemických činnostech Kirsh (1995a, 1996) také často hovoří jako o tzv. **komplementárních činnostech** či **strategiích**, které doplňují vnitřní mentální procesy a jsou vnější komponentou interaktivního kognitivního procesu. Podstatou takových komplementárních činností je podle Kirshe to, že někdy tím nejlepším způsobem, jak vyřešit nějaký kognitivní problém, je přizpůsobit svět svým vlastním schopnostem a nikoli snažit se přizpůsobovat své schopnosti okolí. Tuto strategii Kirsh (1995a) ilustruje na několika z každodenního života velice dobře známých situacích: Bude-li mít například člověk před sebou fotografii nějaké osoby a ta fotografie bude vzhůru nohama, jeho přirozenou reakcí bude to, že vezme fotografii do ruky a otočí ji o 180 stupňů, neboť tvář člověka lze rozpoznat a identifikovat mnohem snáz, když se nachází v obvyklé poloze (viz obrázek 45). Člověk takto k usnadnění percepce a identifikace lidské tváře realizuje činnost, která přizpůsobuje svět jeho vrozeným percepčním schopnostem. Podobně když člověk dostane za úkol naučit se nazpaměť řadu písmen, například RJVZPLKIVZUPHVMCRD, nejdříve bez možnosti se



Obrázek 45: Tvář herce Miroslava Donutila dokáže člověk (znalý českého kulturní prostředí) identifikovat mnohem snáze a rychleji, jestliže si jeho fotografii nejdříve otočí do obvyklé polohy.

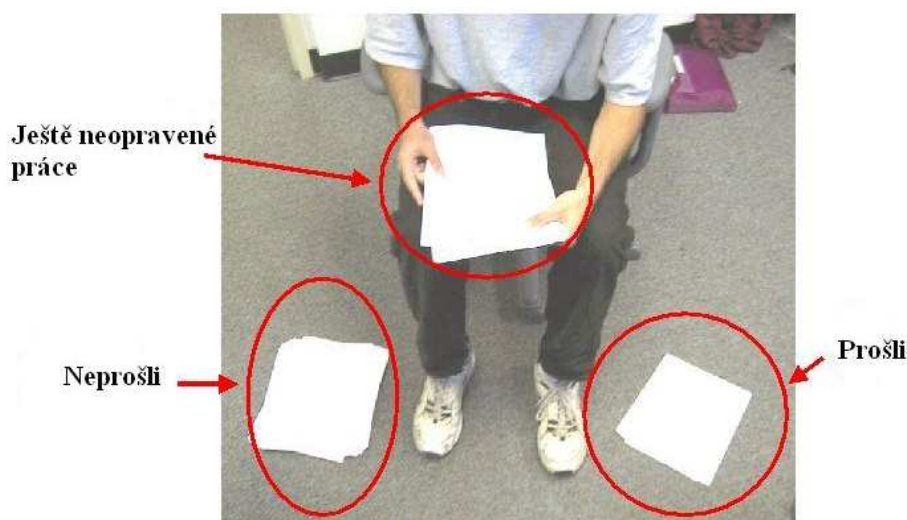
písmen jakkoli dotýkat a manipulovat s nimi a potom s možností si písmena libovolně přeuspořádat, je vysoce pravděpodobné, že v druhém případě brzy objeví způsob, jak přeuspořádat řadu písmen tak, aby mohl zlepšit svůj paměťový výkon. Jedním takovým způsobem by mohlo například přeuspořádání písmen do několika skupinek po třech (RJV ZPL KIV ZUP HVM CRD); jinou možností by bylo jejich uspořádání podle abecedy (CDIJKLMPRRUVVZZ). V případě, že bude mít člověk za úkol určit počet většího množství drobných teček (například), tak člověk celkem přirozeně a bez nějakého velkého přemýšlení šáhne po tužce nebo použije svůj vlastní prst, aby si pomohl s fixováním svého zraku a své pozornosti, a aby tak omezil rušivý vliv sousedních teček. Celý trik spočívá v tom, že člověk nepočítá přímo tečky samotné, ale spíše pohyb tužky či prstu sledujícího jednotlivé tečky, neboť tečky jsou stále příliš drobné na to, aby se daly počítat přímo, ale nejsou příliš drobné na to, aby je nebylo možné sledovat s pomocí tužky nebo prstu. Všem výše uvedeným příkladům komplementárních činností je společné to, že vždy vedou k nějaké změně ve vnějším prostředí, která způsobila změnu povahy problému, se kterým je konfrontována lidská mysl. Kirsh (1995a, s. 212) takto definuje komplementární strategie jako „*jakoukoli organizující činnost, která využívá vnějších struktur k redukci kognitivní zátěže*“. Tyto vnější struktury mohou zahrnovat prsty nebo ruce, tužku, papír, abakus, pravítko, různé symbolické notace, grafy a mnoho dalších prvků, které se teoreticky mohou vyskytovat v bezprostředním okolí člověka. Typickými organizujícími činnostmi jsou pak ukazování, upravování místa a orientace různých předmětů, vypisování různých údajů na papír, manipulace s abakem či s pravítkem nebo s jinými kognitivními artefakty, které dokáží kódovat stav nějakého kognitivního procesu nebo usnadnit percepci informací podstatných z hlediska řešení daného problému. Kognitivní zisky pak většinou spočívají ve vyšší rychlosti řešení problému, ve schopnosti řešit komplexnější problémy, ve snížení pravděpodobnosti chyby při realizaci nějakého plánu činnosti a ve schopnosti se vypořádat s různými rušivými vlivy (distraktory), které jinak stěžují řešení problému.

Hlavní přínos komplementárních strategií tedy spočívá v tom, že umožňují člověku řešit různé typy problémů, které by jinak byly pro kognitivní schopnosti ničím „neozbrojeného“ mozku nezvladatelné. Dennett (2004b, s. 124) takto cituje Dahlboma: „*Stejně jako nelze moc truhlařit jenom holýma rukama, nejde ani moc myslet jenom holým mozkem*.“ Komplementární strategie tohoto efektu dosahují tak, že využívají silných stránek kognitivního profilu lidského mozku, a tím kompenzují jeho slabé stránky. Dobře je tato strategie vidět při hraní scrabblu, kdy lidé různě fyzicky přeuspořádávají jednotlivá písmena, aby vygenerovali co možná nejvíce fragmentárních vstupů, na jejichž základě dokáže

asociační paměť snadno a rychle zkompletovat celá slova, podobně jako dokáže k několika prvním tónům okamžitě dodat zbytek melodie nebo k hlavě kočky zbytek jejího těla skrytého za kusem nějakého nábytku. Jiné komplementární strategie zase využívají schopnosti lidské mysli se dobře orientovat v prostoru. Společným jmenovatelem této skupiny komplementárních strategií je to, že se snaží do určitého prostorového uspořádání fyzikálního prostředí zakódovat informace důležité z hlediska řešení nějakého problému. Výhodou takto zakódovaných informací je to, že jsou pro lidskou mysl velice snadno dostupné; jinak řečeno, „kognitivní náklady“ spojené s jejich získáním jsou mnohem nižší, než kdyby ty samé informace musely být zakódovány v nějaké vnitřní mentální reprezentaci. Takto si například učitel může zakódovat informaci o tom, kteří studenti získali v písemné práci dostatečný počet bodů a kteří nikoli, tím, že si písemné práce jednotlivých studentů uspořádá do dvou hromádek (viz obrázek 47). Tato komplementární strategie „prostorové kategorizace“



Obrázek 46: Při scrabblu hráči často využívají fyzického přeuspořádání jednotlivých písmen v kombinaci se schopnostmi asociační paměti, která i z neúplných vstupů dokáže snadno vygenerovat úplné výstupy v podobě celých slov.



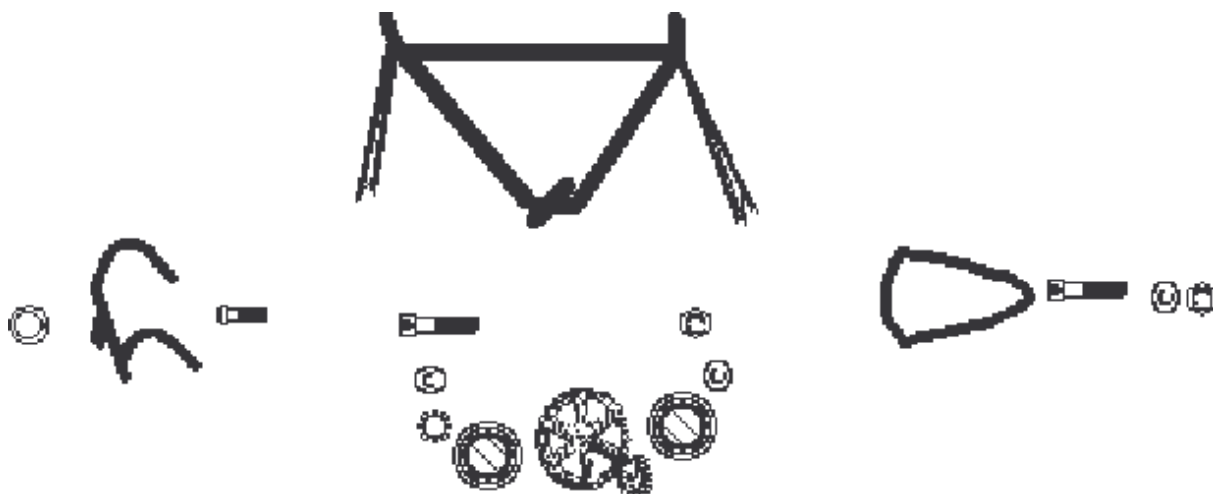
Obrázek 47: Pomocí hromádek lze zakódovat z hlediska řešení nějakého problému důležité informace, které jsou přitom pro lidskou mysl velice snadno a rychle dostupné.

nachází hojného využití také v kancelářích, kde si lidé pomocí hromádek a jejich umístění na stole kódují informace o důležitosti a pořadí jednotlivých úkolů (viz obrázek 48); při praní špinavého prádla si pak zase lidé pomocí hromádek prádla kódují informace o některých jeho vlastnostech, které jsou důležité z hlediska rozhodování o tom, co, jak a s čím vyprat (viz obrázek 48). Do prostorového uspořádání objektů ve vnějším prostředí je také možné



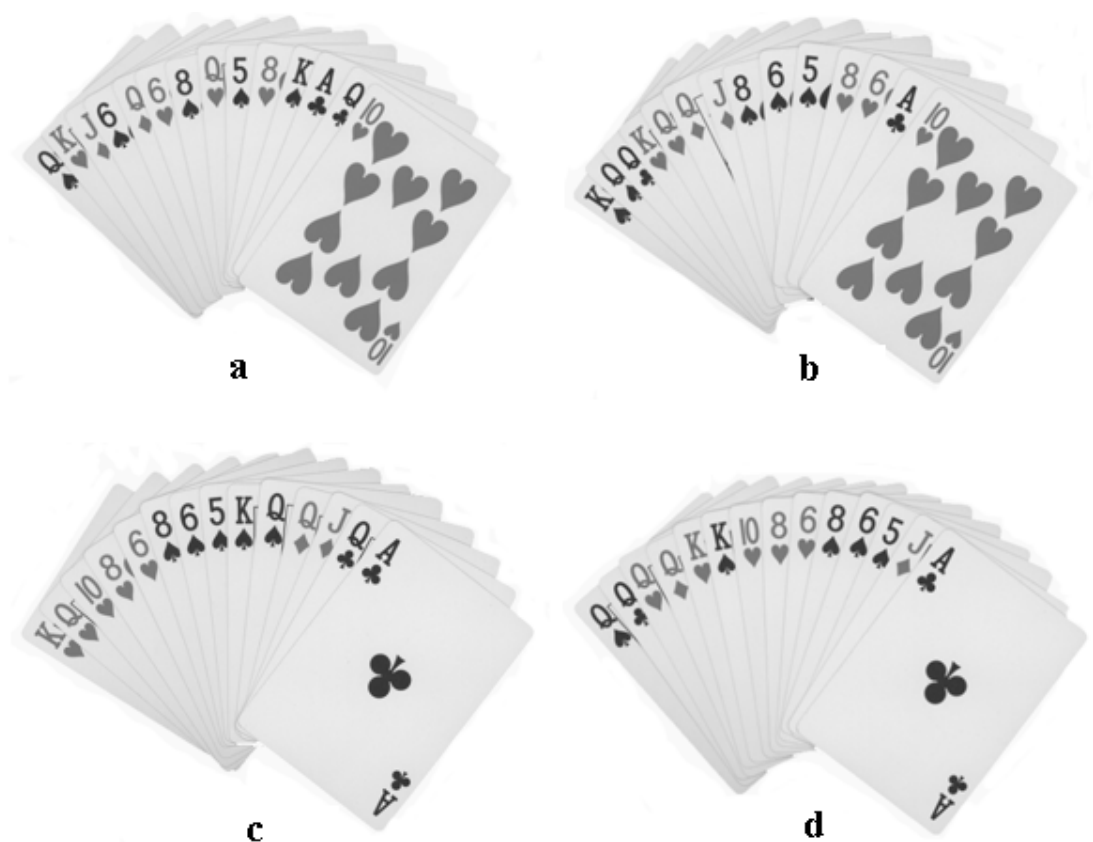
Obrázek 48: V domácnostech i kancelářích lidé často využívají hromádek ke kódování některých informací, které jsou důležité z hlediska řešení některých kognitivních problémů, se kterými se člověk v těchto prostředích setkává.

zakódovat informace klíčové pro řízení komplexních posloupností činností. Typickým příkladem této komplementární strategie je specifický postup při rozebírání nějakého složitějšího technického zařízení, například kola (viz obrázek 49), kdy se jeho jednotlivé součástky kladou do lineárního rozložení v pořadí, v jakém jsou tyto součástky ze stroje



Obrázek 49: Lineární řazení součástek v pořadí, v jakém byly odmontovány, je jednoduchou strategií, jak si usnadnit jejich zpětné smontování dohromady. (převzato z Kirsh, 1995b)

odmontovávány, čímž se značně usnadní opětovné složení stroje dohromady, neboť k tomu nyní člověku postačí zpětně sledovat lineární sled součástí (Kirsh, 1995b). Podobné strategie využívá i většina hráčů karetních her, kteří do prostorového uspořádání karet v průběhu hry neustále kódují informace o svých herních záměrech. Například na obrázku 50 můžeme vidět tři odlišné způsoby uspořádání (b, c, d) stejné sady karet (a), které kódují tři odlišné herní strategie. Z čistě pragmatického hlediska neexistuje žádný rozumný důvod, proč karty takto uspořádávat do skupin, neboť toto seskupování nemá žádný vliv na cíle, kterých se hráč snaží dosáhnout. Toto seskupování má však svůj smysl a význam z hlediska usnadnění některých kognitivních úkolů, které hráč v průběhu hry musí neustále řešit. Jedním z takových úkolů je to, že v průběhu hry si hráč musí v hrubých obrysech vytvořit plán k dosažení některých dílčích cílů hry. A právě tyto plány si hráč může snadno zakódovat do specifického seskupení karet v ruce. Díky tomu, že herní plány jsou takto zevně reprezentovány, hráči si je nemusí pamatovat; mohou je snadno a rychle vyčíst přímo z prostorového uspořádání karet. Navíc s takto uspořádanými kartami hráči dokáží mnohem rychleji posoudit, zda mu posuzovaná karta může posloužit ke zkompletování série; hráči se také díky tomuto způsobu kódování při hře dopouštějí menšího počtu chyb (Kirsh, 1996).



Obrázek 50: Tři způsoby prostorového uspořádání (a, c, d) jedné sady karet (b), které kódují tři odlišné herní strategie. (převzato z Kirsh, 1996)

Shrneme-li vše výše uvedené, můžeme konstatovat, že způsob, jakým lidská mysl obvykle řeší nejrůznější problémy, neodpovídá tradičnímu (symbolovému) modelu plánování a řešení problémů jako úplné vnitřní (mentální) specifikace posloupnosti činností, které musí být pouze zrealizovány ve vnějším světě, jehož jedinou funkcí je poskytovat lidské mysli vstupy specifikující danou problémovou situaci a umožňovat jí realizovat její předem promyšlené plány řešení problémů. Ve skutečnosti lidé využívají celou škálu (komplementárních) strategií, které jako nedílnou součást procesu řešení problému zahrnují různé manipulace s objekty ve vnějším prostředí. Intelligence člověka pak spočívá především v jeho schopnosti vhodným způsobem strukturovat prostředí a manipulovat objekty v něm tak, aby se změnila povaha kognitivních operací, které musí lidský mozek (vyznačující se kognitivním profilem se specifickým vzorcem silných a slabých stránek) provádět v průběhu řešení různých problémů. To mimo jiné znamená, že vnější prostředí není v žádném případě striktně odděleno od vnitřního prostředí lidské mysli (jak to sugerují tradiční symbolové modely lidské kognice); naopak, vnější prostředí hraje aktivní roli v řízení a v „tvarování“ činnosti lidského mozku, takže ten je pak schopen i se svým relativně omezeným kognitivním profilem plnit funkce, které by jinak ležely zcela mimo dosah jeho základních schopností.

Toto podpůrné prostředí přitom nemusí mít čistě jenom fyzikální povahu, ale může se také jednat o sociokulturní prostředí: Podobně jako existující dopravní síť silnic usměrňuje a svádí celkový tok dopravy určitými specifickými místy, usměrňuje aktivitu a činnost jednotlivých myslí také struktura sociálních vztahů, institucí či organizací v podobě různých zvyků, norem, směrnic, pravidel, hodnot apod. Jednoduchým příkladem takové kulturní normy je fronta – kulturní zvyk, který významným způsobem snižuje nároky kladené na lidskou paměť v sociální interakci tím, že „nutí“ lidi používat jejich vlastní těla k vytvoření prostorové paměti, která kóduje pořadí, v jakém jednotliví klienti či zákazníci přicházejí ke zdroji nějakého statku a v jakém (spravedlivém pořadí) by tedy také měli mít k danému statku přístup (viz obrázek 51). Komplexnějším příkladem takového sociálně usměrňovaného procesu řešení problému by byla zde již jednou zmiňovaná lodní navigace (viz s. 39), kde každý člen navigačního týmu v interakci s nějakým kognitivním artefaktem (např. s námořním logaritmickým pravítkem) tvoří nějaký lokální distribuovaný kognitivní systém plnící určitou specifickou kognitivní funkci a kde jsou prostřednictvím různých pracovních norem, směrnic a zvyků tyto lokální distribuované kognitivní systémy koordinovány tak, aby vytvořili globální distribuovaný kognitivní systém, který jako celek dokáže vyřešit to, co je jinak zcela mimo dosah jeho jednotlivých funkčních částí. Do stejné kategorie by patřila také

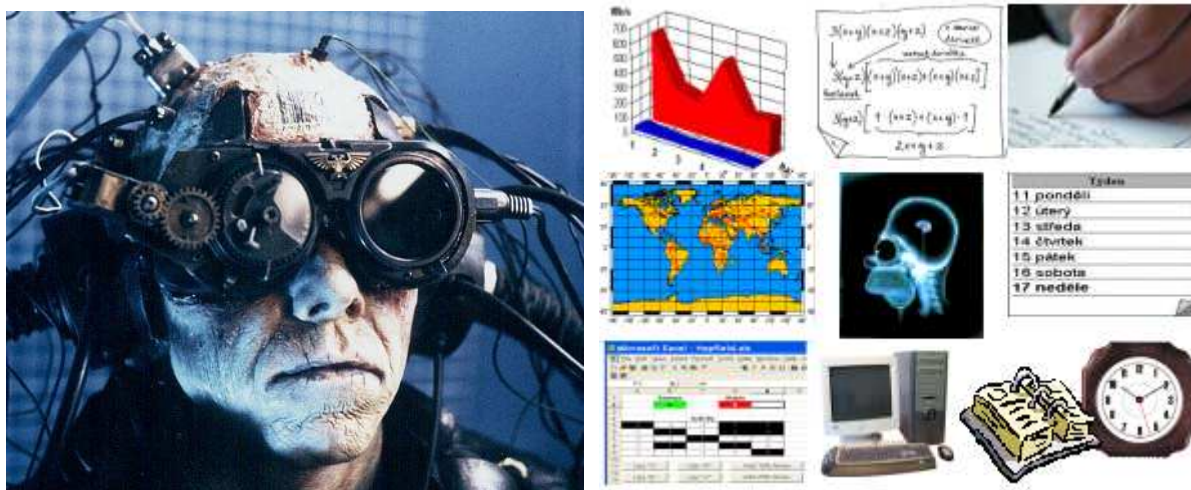


Obrázek 51: Fronta je kulturní norma či zvyk, který umožňuje vytvořit prostorovou paměť pořadí příchodu jednotlivých zákazníků. Ti používají svá vlastní těla, resp. jejich relativní umístění v prostoru tvořeném těly dalších zákazníků, ke kódování pořadí, v jakém jednotliví zákazníci přicházejí. (převzato z Hutchins, 2004, s. 22)

manufakturní výroba různých navigačních a logaritmických tabulek „sociálními počítači“ 19. století, tvořenými koordinovanou činností hierarchie lidských počítačů, které prováděli pouze rutinní a jednoduché výpočty a své výsledky odesílali k dalšímu zpracování dalším lidským počítačům, aniž by přitom měli sebemenší představu o fungování celého výpočetního soukolí (viz s. 40). Podobným způsobem usměrňuje chování jednotlivých myslí také širší sociokulturní prostředí firmy nebo podniku. Vzhledem k tomu, že firmy „žijí“ ve vysoce konkurenčním prostředí, existuje zde permanentní selekční tlak na to, aby se firmy chovaly racionálně a efektivně, neboť pouze firmy, které se chovají tímto způsobem, mohou dlouhodobě přežít a prosperovat. Důsledkem těchto selekčních tlaků je to, že většina existujících firem inkorporovala takové vnitřní regulační mechanismy v podobě různých směrnic, norem a hodnot, které nutí její zaměstnance chovat se způsobem, který na globální úrovni vede k tomu, že se podnik jako celek chová racionálně a efektivně. Clark (1998b) tímto mechanismem vysvětluje, jak je možné, že se velké sociální organizace (jako jsou firmy nebo státy) řídí principy racionálního myšlení (a že tedy lze na jejich chování úspěšně aplikovat tradiční ekonomické modely rozhodování vycházející z předpokladu neomezené racionality *Homo oeconomicus*), přestože se tyto velké sociální organizace skládají z lidí, kteří disponují pouze omezenou racionalitou (Simon, 1982).

Podle Clarka (1997, s. 191) obecně platí, že specificky lidská inteligence (*advanced reason*) je funkcí mozku „podepíraného“ komplexní strukturou vnějšího prostředí (*scaffolded*

brain), tj. „mozku v jeho tělesném kontextu, interagujícího s komplexním světem fyzikálních a sociálních struktur. Tyto vnější struktury usměrňují a rozšiřují problémy řešící aktivity...mozku, jehož role z velké části spočívá v zajišťování série opakovaných, lokálních a



Obrázek 52: Dvě různá pojetí kyborga. Podle Andyho Clarka (2003) nejsou kyborgové žádnou sci-fi, ale holou skutečností už přinejmenším po dobu posledních 50 000 let, kdy lidé k vylepšení svého kognitivního profilu využívají řady různých externích kognitivních pomůcek a nástrojů (včetně jazyka), které jsou součástí jejich kulturního a technologického prostředí. Těsné propojení (funkce) biologického mozku s těmito vnějšími kognitivními technologiemi přitom není vůbec závislé na jejich přímé fyzické implantaci do lidského těla, neboť „to, na čem [z kognitivního hlediska] opravdu záleží, [je] pouze hladká a bezproblémová integrace člověka a stroje a výsledná transformace našich schopností, projektů a životních stylů“¹ (Clark, 2003, s. 24). Student píše diplomovou práci, který k organizaci a realizaci své práce využívá počítač, kalkulačku, hodinky, plánovací kalendář, různá nebiologická úložiště informací, formální systémy, symbolické notace, grafy, schémata, mapy, experimentální a statistické postupy a obecně poznatky a myšlenky druhých lidí... tak podle Clarka není o nic méně kyborgem než tradiční hybrid člověk-stroj, jehož tělo je prošípováno spoustou drátů, čipů, kamer a dalších podobných součástek. „Tyto různé nástroje a pomůcky generují informace, nebo je ukládají, nebo je transformují... [a] tím ovlivňují [studentovu] schopnost řešit problémy stejně dramatickým způsobem jako různé softwarové balíčky ovlivňují výkon jednoduchého PC.“² (Clark, 1998a, s. 162) V podobném duchu francouzský filozof Jean Baudrillard konstatuje, že „reálně a subjektivně jsem člověk, virtuálně a prakticky jsem stroj. [...] Avšak virtuální stroje a technologie mě neodcizují. Tvoří spolu se mnou integrovaný obvod. [...] Velké i malé počítače, televize, video, dokonce i fotoaparát jsou jako kontaktní čočky, průhledné protézy, které jsou do těla integrovány natolik, že k němu již patří téměř geneticky, jako například kardiostimulátor...“ (citováno podle Kelemen, 1995, s. 55-56)

¹ „... what really matters might be just the fluidity of the human-machine integration and the resulting transformation of our capacities, projects, and lifestyles.”

² „These various tools and props thus act to generate information, or to store it, or to transform it, or some combination of the three. In so doing, they impact our individual and collective problem-solving capacities in much the same dramatic ways as various software packages impact the performance of a simple pc.”

vzorci kompletujících reakcí.“¹ Lidský mozek překračuje výpočetní možnosti své architektury tím, že ve vnějším prostředí vytváří (ale hlavně po svých bezejmenných předcích „dědí“) různé podpůrné fyzikální a sociální struktury, které mu pomáhají sestavit a zkoordinovat komplexní posloupnosti poměrně snadno zvládnutelných „epizod“ řešení dílčích problémů. Podpůrné struktury vnějšího prostředí takto mozku umožňují použít jeho vlastní omezený kognitivní profil kompletovače vzorců ke zvládnutí problémových domén, které jsou jinak pro „holý“ mozek zcela mimo jeho dosah. **Individuální lidská mysl, resp. mozek je podle tohoto pojetí pouze jedním malým „kolečkem“ ve velkém bio-socio-kulturně-technologickém stroji.** Teprve interakce mezi biologickými, sociokulturními a technologickými komponentami dává vyvstat vyšším kognitivním funkcím jako je usuzování, řešení problémů apod. **Chyba tradiční kognitivní vědy spočívá v tom, že se snaží vtěsnat všechno to, co tento bio-socio-kulturně-technologický stroj jako celek dokáže, do individuální lidské mysli, resp. do mozku.** Teorie distribuované kognice proto navrhuje rozšířit předmět kognitivní vědy za hranice lidského mozku a těla a dívat se na lidskou inteligenci jako na druh *rozšířeného fenotypu (extended phenotype)*².

¹ „the brain in its bodily context, interacting with a complex world of physical and social structures. These external structures both constrain and augment the problem-solving activities of the basic brain, whose role is largely to support a succession of iterated, local, pattern-completing responses.“

² Zrzavý, Storch a Mihulka (2004, s. 71-72) charakterizují rozšířený fenotyp následujícími slovy: „Fenotypové projevy genu chápeme jako souhrn jeho účinků na organismus. Je jistě možné, že projevy určitého genu budou omezeny vždy jen na těla organismů, ve kterých tento gen přebývá. To je však pouze jedna z možností. Fenotypovým projevem genu je totiž souhrn všech účinků, které má tento gen na vše, co jej obklopuje. Fenotyp genu představují nástroje, které dotýčný gen pomáhají protlačit do dalších generací, a tyto nástroje mohou přesahovat vlastní tělesnou schránku organismu, ve kterém gen právě dlí. Pěkným příkladem této situace mohou být bobří hráze, ptačí hnízda, pavučiny, schránky larev chrostků nebo řekněme motocykl značky Harley Davidson.“

2 Konekcionistické modelování kognitivních procesů

„The brain exemplifies the fact that it is all right to be very stupid if you're well connected.“

DAN LLOYD, *Simple Minds*

„The large brain, like large government, may not be able to do simple things in a simple way.“

DONALD HEBB, *The Organization of Behavior*

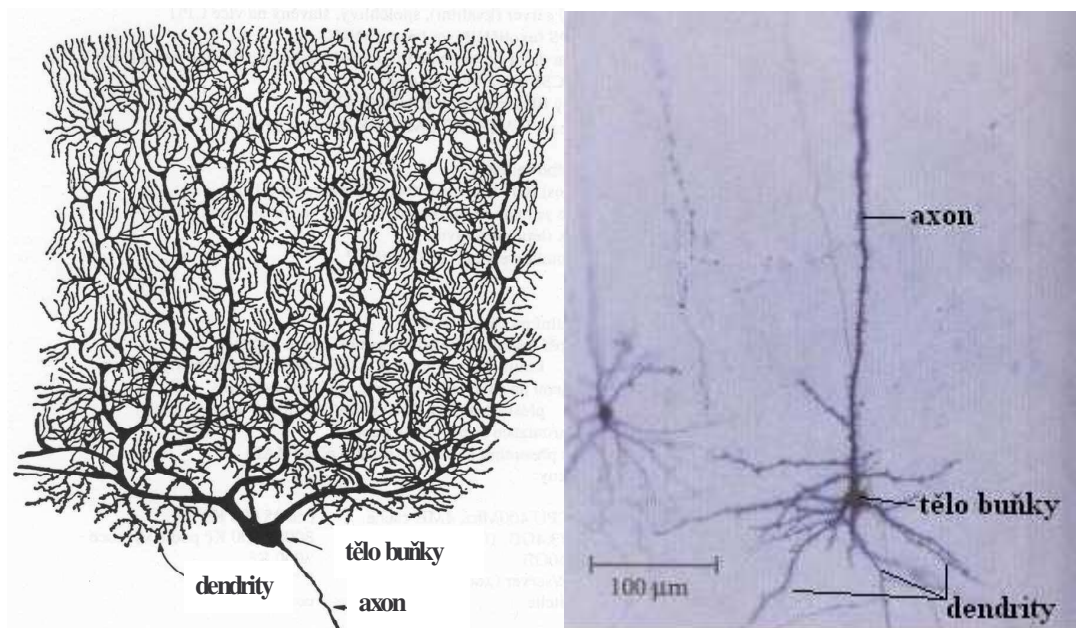
„Představme si, že jsme cestovatelé objevující novou zemi. Po příplutí k pobřeží vidíme úzkou zátoku s jemným pískem a pásem kokosových palm. Když začneme chodit po ostrově, vynořují se před námi stále nové a nové skutečnosti: vrchy, močály, lesy, skalní útesy apod. Jak se v této krajině orientovat? [...] První věc, jež nás... napadne, je vyrobit si mapu. [...] Dobře zhotovená mapa totiž přehledně znázorní hlavní prvky krajiny (objekty), jejich rozměry, velikost (atributy), ale i vzdálenost mezi nimi (vztahy mezi objekty). Samozřejmě, že na mapě nenajdeme všechny vrchy a vršky, nebo dokonce všechny stromy: to by potom nebyla mapa, ale úplná kopie – zdvojení reality. To, co jsme udělali, byl výběr základních prvků systému s jejich základními vztahy. Mapa je modelem, který reprezentuje danou krajinu.“

JÁN FERJENČÍK, *Úvod do metodologie psychologického výzkumu. Jak zkoumat lidskou duši*

2.1 Principy zpracovávání informací v lidském mozku

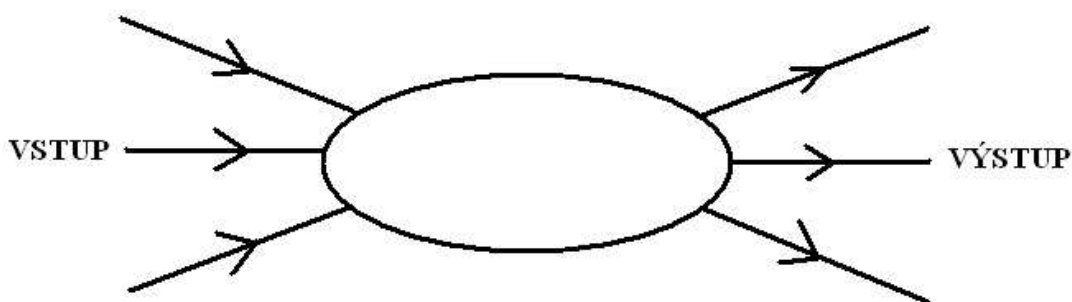
Výpočetní architektura konekcionistických modelů (umělých neuronových sítí, PDP systémů) se přímo inspirovuje mechanismem zpracovávání informací v lidském mozku, jehož výpočetní činnost lze na určité úrovni abstrakce a při určitém zjednodušení charakterizovat prostřednictvím následujících pěti principů (McLeod, Plunkett, Rolls, 1998):

1) *Neurony integrují informace.* V mozku se nachází mnoho různých druhů neuronů, jejichž stavba se může v různých ohledech značně lišit (viz obrázek 53). Přes tyto rozdíly plní všechny neurony stejnou funkci, která spočívá a) v integrování příchozích informací o míře aktivity jednoho souboru neuronů a b) v odesílání informací o aktivitě této skupiny neuronů jinému souboru neuronů. Ve značně zjednodušeném a zidealizovaném konceptuálním modelu skutečného biologického neuronu, který zachycuje podstatu jeho funkce (tzv. *klasický neuron*) se tento proces odehrává ve třech na sebe navazujících stádiích: Neuron nejdříve na svých dendritech přijímá excitační nebo inhibiční signály od jiných neuronů prostřednictvím synaptických spojů. Jestliže suma přijatých signálů překročí určitou prahovou hodnotu, neuron sám začne být aktivní a informaci (signál) o tom předá prostřednictvím svého axonálního výběžku dalším neuronům, které tento signál přijmou na svém dendritu jako jeden z mnoha svých vstupů. V konekcionistickém modelu je funkční role neuronu implementována



Obrázek 53: Neurony. Nalevo je Purkyněho buňka, která se nachází v lidské mozečku; napravo je pak neuron nacházející se v primární zrakové kůře. Přestože se struktura obou druhů nervových buněk značně liší, funkce, kterou plní, je u obou buněk totožná: Obě buňky na svých dendritech přijímají vstupní informace v podobě excitačních nebo inhibičních signálů od dalších neuronů. Tyto signály putují do těla buňky, kde jsou jednotlivé signály integrovány. Jestliže suma těchto signálů překročí určitou kritickou hodnotu, neuron po svém axonálním výběžku vyšle signál směrem k dalším neuronům, se kterými je propojen.

v podobě jednoduché výpočetní jednotky, která přijímá signály od dalších výpočetních jednotek prostřednictvím pozitivních nebo negativních spojů; signály přicházející po těchto vstupních spojích mohou zvyšovat nebo naopak snižovat míru aktivity centrální jednotky, která přicházející signály sumuje. Informaci o celkové sumě přijatých signálů nakonec výpočetní jednotka odesílá dalším výpočetním jednotkám, se kterými je propojena. Výpočetní jednotka konekcionistického modelu tak plní stejnou funkční roli jako biologický

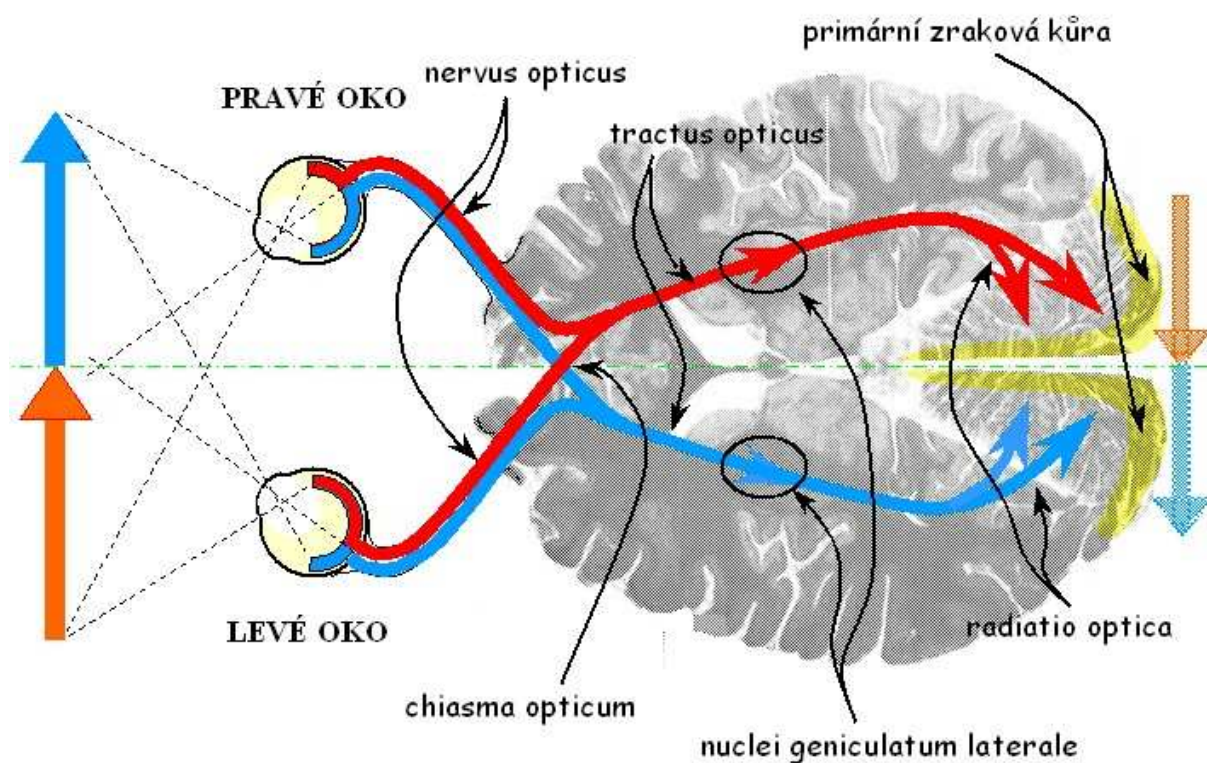


Obrázek 54: Výpočetní jednotka konekcionistického modelu představuje funkční ekvivalent biologického neuronu: Jestliže suma vstupů překročí určitou prahovou hodnotu, výpočetní jednotka vyšle signál k dalším výpočetním jednotkám, se kterými je propojena. (s úpravami převzato z McLeod, Plunkett, Rolls, 1998, s. 12)

(resp. klasický) neuron, neboť to, co dělá, je to, že jedné skupině výpočetních jednotek předává informace o míře aktivity jiného souboru výpočetních jednotek.

2) *Neurony předávají informace o míře přijatých vstupních informací.* Neurony nepředávají svou aktivitou informace pouze o tom, že na svém vstupu přijali nějaké signály, ale informují ostatní neurony také o tom, kolik těchto přijatých signálů bylo, resp. jak velká je míra jejich stimulace. Výstupní informace neuronu je takto vždy v nějakém zákonitém vztahu (ne nutně lineárním) k úrovni vlastní stimulace. Biologické neurony míru své stimulace kódují v podobě frekvence signálů, které vyšlou za určitou časovou jednotku k dalším neuronům, přičemž platí, že čím vyšší úroveň stimulace, tím vyšší frekvence vzruchů (signálů). V konekcionistických modelech není informace o úrovni přijatých signálů kódována v podobě frekvence vzruchů, ale prostřednictvím jednoduché číselné hodnoty, tzv. míry aktivace.

3) *Struktura mozku má vrstevnatou povahu.* V lidském mozku tvoří nervové buňky dva základní typy fyzicky relativně nezávislých struktur (tzv. *lokální obvody*) - *jádra* a *vrstvy*. Příkladem jader jsou *bazální ganglia* (důležitá součást funkčního systému zodpovědného za motorické funkce) či *amygdala* (součást funkčního systému emotivity); ve vrstvách jsou pak uspořádány především jednotlivé korové oblasti, jejichž tkáň na většině míst tvoří šest vrstev nervových buněk (Koukolík, 2005; Damasio, 2000). Zpracovávání informací v lidském mozku probíhá postupným šířením (mapováním) vzorců nervové aktivity (*neuronálních map/reprezentací*) z jedné struktury do druhé. Například při motorické reakci na nějaký vizuální podnět projde vzorec aktivace vzniklý dopadem světla na sítnici celou řadou nervových struktur, které výchozí aktivační vzorec fotoreceptorů postupně přetransformují do podoby aktivačního vzorce neuronů motorické kůry. Na obrázku 55 je zobrazena část této dráhy: Světlo dopadá na fotoreceptory sítnice, které energii fotonů přemění na elektrochemické signály nervové soustavy. Takto vzniklý vzorec aktivace je optickým nervem (*nervus opticus*) přenesen do zrakového thalamu (*nuclei geniculatum laterale*) a odtud pak nervovými vlákny (*radiatio optica*) do primární zrakové kůry (V1, Brodmannova area 17). Odtud jsou vzorce nervové aktivity přeneseny do dalších korových oblastí zajišťujících jejich další zpracování (V2, V3...) a nakonec také určitou motorickou odpověď. Každá z pomyslných zastávek na cestě od zahlédnutí nějakého objektu, například hada, k nějaké motorické reakci typu boj/útěk (*nuclei geniculatum laterale*, zraková kůra, motorická kůra...) představuje seskupení miliónů navzájem hustě propojených neuronů, které transformují přicházející vzorce nervové aktivity a takto transformované aktivační vzorce

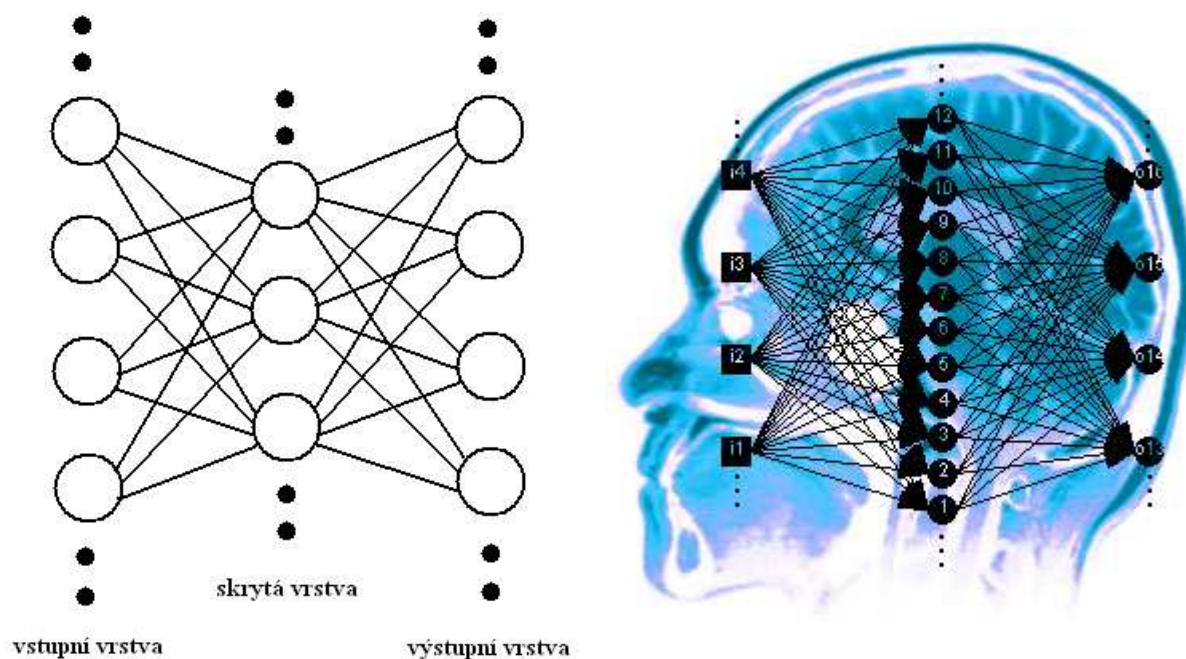


Obrázek 55: Schematické znázornění některých nervových drah a center, která se spolupodílejí na zpracovávání zrakových podnětů. (s mírnými úpravami převzato z ppt prezentace Ivana M. Havla k přednáškám „Přirozené a umělé myšlení“ konaným ve školním roce 2005/2006 na pracovišti Centra Teoretických Studií v Praze)

pak přeposílají jako vstup do další vrstvy neuronů.¹ Tento způsob organizace procesu zpracovávání informací je vtělen i do konekcionistických modelů, které mají většinou podobu vícevrstevných neuronových sítí (*multilayered neural networks*), které se skládají z jedné vstupní vrstvy (*input layer*), kde jsou sítě prezentovány vstupní podněty, z jedné nebo více skrytých vrstev (*hidden layers*), které jsou propojeny se vstupní vrstvou a s vrstvou výstupní, která kóduje odpověď neuronové sítě na prezentovaný vstupní podnět. Aktivační vzorce se v takovém typu neuronové sítě šíří většinou jedním směrem a to od vstupní vrstvy přes vrstvu skrytou k vrstvě výstupní; jednotlivé vrstvy neuronové sítě však lze propojit i složitějším způsobem a umožnit například výstupní vrstvě zpětnovazebně ovlivňovat činnost některé ze skrytých vrstev. V průběhu šíření aktivačních vzorců mezi jednotlivými vrstvami jednotek dochází k postupné transformaci vstupních vzorců aktivace, které mohou mít na konci celého procesu zpracovávání informací podobu například správné výslovnosti

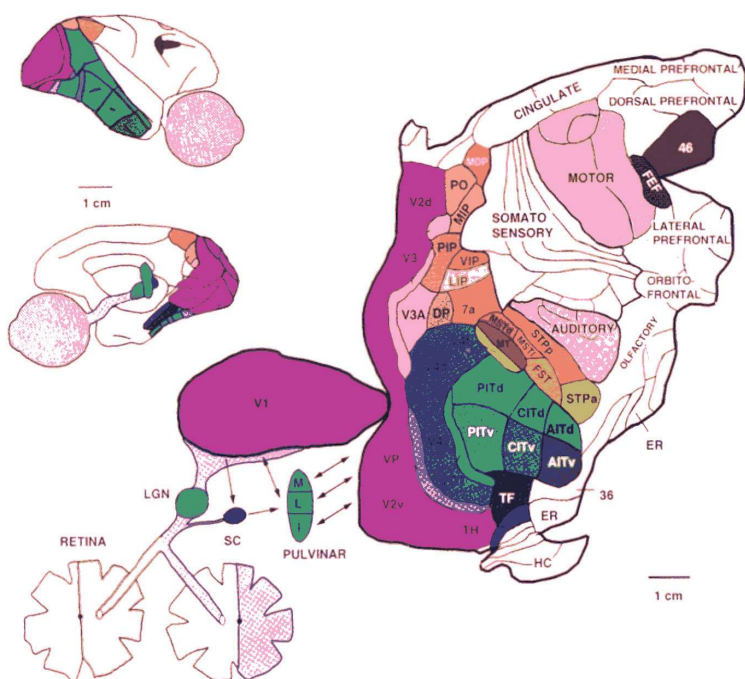
¹ Podobně jako u produkčního systému (viz s. 77-79) i zde lze jako ilustraci celého procesu zpracovávání informací použít jednoduchý slovní hlavolam („slovní řetěz“) spočívající ve změně jednoho slova na druhé prostřednictvím postupné obměny jednotlivých hlásek, s tím omezením, že všechna slova, která vzniknou v každém jednotlivém kroku obměny jedné hlásky, musí být smysluplným podstatným jménem v 1. pádě jednotného čísla: Útočné chování vygenerované lidským mozkem v reakci na hada je tak možné metaforicky vidět jako postupnou transformaci následující sady slov: **had** – hod – rod – rok – rek – rej – roj – **boj**.

přečteného slova, gramaticky správně zformulované věty, potvrzení pravosti podpisu atd. Celé to trochu připomíná situaci, kdy si občan na úřad podá dotaz, který projde rukama nepřehledné houštiny hierarchie velkého počtu úředníků, z nichž se každý k dotazu nějak vyjádří, trochu přeformuluje odpověď ostatních úředníků..., ale nakonec se člověk navzdory všemu tomu chaosu a zmatku přece jenom dočká rozumné a stručné odpovědi v podobě úhledně napsaného dopisu.



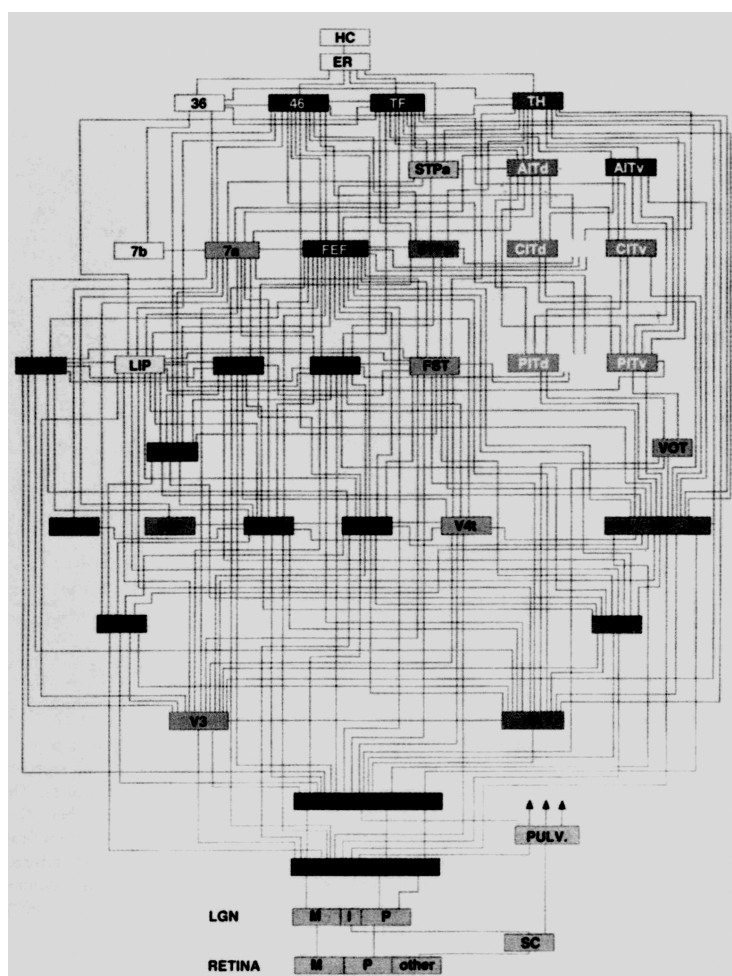
Obrázek 56: Vícevrstevná neuronová síť představuje konekcionistický model zpracovávání informací v lidském mozku prostřednictvím postupného šíření (mapování) vzorců nervové aktivity z jedné nervové struktury do druhé. (levý obrázek převzat s úpravami z McLeod, Plunkett, Rolls, 1998, s. 13)

Zpracovávání informací v lidském mozku je ve skutečnosti mnohem komplikovanější než jak to sugeruje představa aktivačních vzorců, které se postupně šíří a které jsou postupně transformovány v hierarchicky uspořádaných vrstvách jednotlivých neuronálních struktur (tzv. *bottom-up processing*, zpracovávání informací „odspodu-vzhůru“). Tento hierarchicky uhlazený popis struktury mozku je pouze velice hrubým přiblížením skutečné komplexní konektivity, kterou můžeme nalézt v mozku člověka nebo jakéhokoli jiného vyššího savce. Takto se například u makaka prokázalo (Felleman, Van Essen, 1991), že mezi 32 jednotlivými zrakovými korovými oblastmi, ze kterých se zrakový kortex makaka skládá (viz obrázek 57), existuje hustá síť reciprokých (zpětnovazebních) spojů, které vedou jak směrem „odspodu-vzhůru“ (například z V1 do V2) tak i směrem „shora-dolů“ (například z V2 do V1) nebo „do stran“ (například mezi dvěma podoblastmi V1) (viz obrázek 58). Tuto komplexní

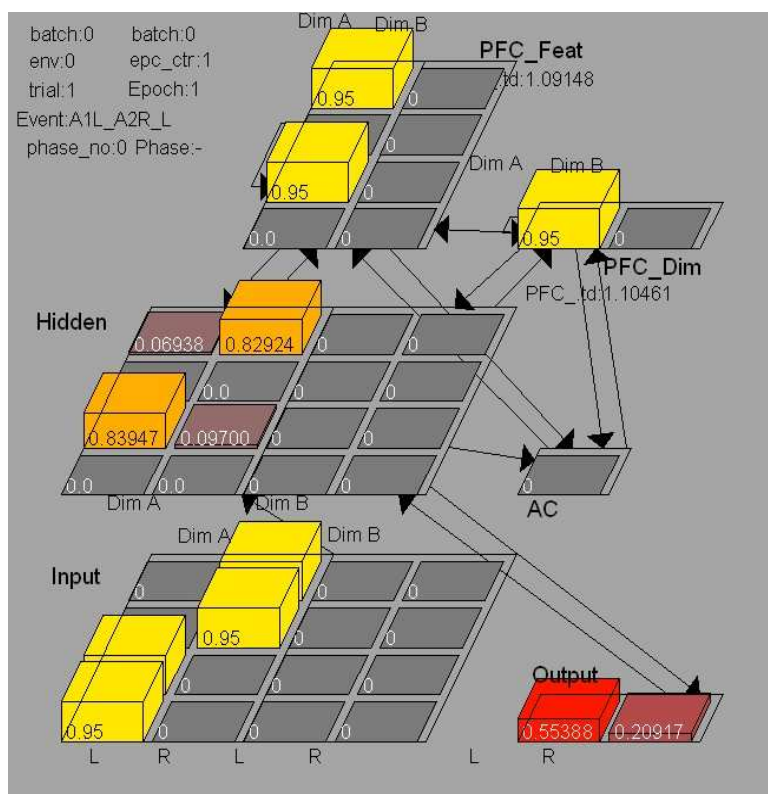


konektivitu biologických neuronových sítí se snaží - více či méně úspěšně - zachytit biologicky více plausibilní konekcionistické modely druhé a třetí generace (viz obrázek 59).

Obrázek 57: Mapa 32 zrakových korových oblastí v mozku makaka.



Obrázek 58: Diagram konektivity (wiring diagram) zrakové kůry makaka.



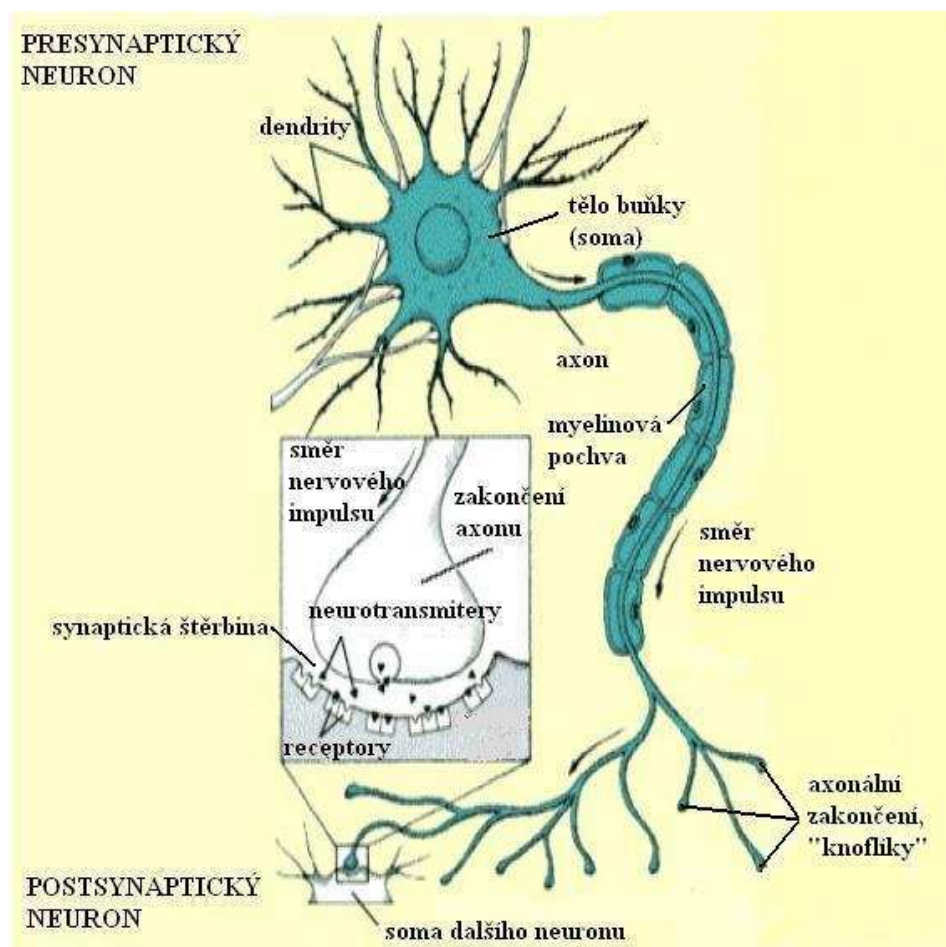
Obrázek 59: Příklad umělé neuronové sítě implementující biologicky plausibilnější strukturu konektivity, která zahrnuje četné zpětnovazební spoje propojující jednotlivé vrstvy výpočetních jednotek jak směrem „odspodu-vzhůru“ tak i směrem „shora-dolů“. Jedná se o konekcionistický model simulující funkci prefrontálního kortexu podílejícího se na řešení dynamické kategorizační úlohy - zjednodušené verze Wisconsinského testu třídění karet, ve které je probandovým úkolem třídit prezentovaný podnětový materiál na základě jednoho ze dvou třídících/kategorizačních kritérií (dimenzí), ve kterých podnětový

materiál může variovat; vzhledem k tomu, že kategorizační kritérium se v průběhu testu mění, je proband nucen aktuálně používané kategorizační kritérium vyvozovat na základě zpětné vazby, které se mu dostává od administrátora. Zdraví probandi dokáží relativně rychle přepínat mezi jednotlivými třídícími pravidly, zatímco probandi s poškozeným prefrontálním kortexem mají tendenci perseverovat a setrvávat u stávajícího klasifikačního kritéria navzdory poskytované zpětné vazbě o chybném výkonu. Více informací k samotnému modelu viz O'Reilly, Noelle, Braver, Cohen (2002). (vytvořeno s pomocí simulačního softwaru PDP++ na základě cvičného příkladu z knihy Randalla O'Reillyho a Yuko Munakatové (2000))

4) Vliv jednoho neuronu na druhý závisí na síle spoje mezi nimi. V lidském mozku může být jeden neuron propojen s $3-10 \times 10^3$ dalších neuronů, tzn. že jeden průměrný neuron v lidském mozku může ovlivňovat činnost okolo 5 tisíc dalších neuronů. Místem, kde k tomuto ovlivňování dochází, je synapse, která je tvořena zakončením axonálního výběžku vysílajícího neuronu (tzv. presynaptická část synapse) a počátkem dendritického výběžku přijímajícího neuronu (tzv. postsynaptická část synapse): Vzruch (nervový impuls, signál) přivedený obalem axonu¹ presynaptického neuronu uvolní z jeho zakončení (tzv.

¹ Tělo a výběžky neuronu jsou kryty povrchovým obalem (plazmatickou membránou), na které lze změřit (v případě klidového stavu neuronu) stálé elektrické napětí, tzv. klidový membránový potenciál, který je důsledkem toho, že neurony vytvářejí a udržují (pomocí iontových pump) rozdílné koncentrace iontů (elektricky nabitých částic: K^+ , Na^+ , Ca^{2+} a Cl^-) na obou stranách plazmatické membrány, čímž vzniká napěťový rozdíl mezi vnější a vnitřní stranou membrány; vnitřek buněk je přitom negativní vůči vnějšku. Klidový membránový potenciál dosahuje v různých buňkách hodnot -9 až -100 mV, v neuronech pak -40 až -90 mV, většinou však -70 mV. Vzruchy (signály) jsou pak vlny depolarizace (povrchové membrány neuronů), prudkého a rychlého

synaptických váček) molekuly neurotransmiteru, nervového přenašeče; ty projdou synaptickou štěrbinou a naváží se na své receptory v membráně druhého (postsynaptického) neuronu, přičemž vazba neurotransmiteru na receptor ovlivní chování neuronu ve smyslu



Obrázek 60: Schéma přenosu elektrochemických signálů mezi neurony.

zvýšení (excitace), nebo naopak snížení pravděpodobnosti, že tento neuron začne vysílat signály k dalším neuronům (inhibice). Míra vlivu, který má jeden neuron na druhý (ať už ve smyslu excitace nebo inhibice jeho aktivity), je determinována silou či efektivitou synaptického spoje mezi nimi. Na efektivitu synaptického spoje může mít vliv množství

poklesu (elektrického) napětí (mezi vnitřkem a vnějškem neuronu). Vlna této depolarizace se šíří obalem neuronu směrem k synaptickému zakončení, kde dochází k uvolnění neurotransmiterů ze synaptických váček do synaptické štěrbin. Uvolněné neurotransmitery difundují směrem k membráně postsynaptického neuronu, na které se naváží na své receptory. To má za následek změnu propustnosti povrchu postsynaptického neuronu, která se projeví buď zvýšením (excitace), nebo naopak snížením pravděpodobnosti jeho aktivace (inhibice). Inhibiční účinky v mozkové kůře zprostředkovává především aminokyselinový neurotransmiter *GABA* - kyselina *gamaaminomáselná*, která po navázání na své receptory na postsynaptickém neuronu otevírá chloridové kanály, kterými dovnitř postsynaptického neuronu mohou vnikat chloridové aniony; excitační účinky pak v mozkové kůře zprostředkovává především aminokyselinový neurotransmiter *glutamát*, který po navázání na své receptory v membráně postsynaptického neuronu otevírá sodíkové kanály, kterými dovnitř neuronu mohou vnikat sodíkové kationy. (Sternberg, 2002; Fišar, Jirák, 2001)

neurotransmiterů, které je k dispozici, počet a citlivost receptorů na straně postsynaptického neuronu, efektivita re-uptakeového mechanismu vychytávajícího ze synaptické štěrby nevyužité neurotransmitery k dalšímu použití apod. Informace o síle či efektivitě spojů mezi výpočetními jednotkami je reprezentována i v rámci konekcionistických modelů, kde se nazývá **váha spoje**.

5) *Učení je dosahováno prostřednictvím změny síly spojů mezi neurony.* Jak již bylo výše uvedeno, v lidském mozku dochází ke zpracovávání informací prostřednictvím šíření aktivačních vzorců mezi jednotlivými strukturami neuronů, které postupně transformují vstupní vzorce aktivace na výstupní vzorce aktivace. Na způsob šíření aktivačních vzorců mezi jednotlivými neurony mají přitom největší vliv síly (efektivita) spojů mezi jednotlivými neurony. Tzn. že především na nich závisí, zda mozek dokáže na nějaký podnět odpovídajícím způsobem zareagovat. A jestliže učení znamená schopnost na základě svých zkušeností modifikovat a adaptovat vzorce svých behaviorálních reakcí na určitou skupinu podnětů, je pak jen logické předpokládat, že učení v lidském mozku bude implementováno prostřednictvím nějakého mechanismu, který bude schopen systematicky a v souladu se získanými zkušenostmi měnit sílu synaptických spojů mezi neurony. A skutečně je dnes již celkem spolehlivě prokázáno, že učení v lidském mozku (kromě jiného)¹ zahrnuje změnu síly synaptických spojů mezi neurony - tzv. *dlouhodobá potenciace a inhibice (long-term potentiation/inhibition)*. „*Ať už se učíte chodit, řešit diferenciální rovnice či číst francouzskou poezii ze šestnáctého století, ve vašem mozku se objevuje pokaždé to samé – pozměňování synapsí.*“ (Cumminsová, 1998, s. 40) Také v konekcionistických modelech je učení implementováno prostřednictvím mechanismů – tzv. učících algoritmů -, které mění váhy spojů mezi jednotlivými výpočetními jednotkami v souladu se zpětnou vazbou, kterou neuronová síť dostává z vnějšího prostředí. V typickém případě začíná neuronová síť s náhodně nastavenými vahami spojů, takže náhodná je také odpověď, kterou neuronová síť dává na své výstupní vrstvě na daný podnět; učící algoritmus automaticky počítá velikost rozdílu mezi aktuálním a požadovaným výstupem a tuto informaci používá k výpočtu mírné úpravy vah spojů, tak aby se výkon neuronové sítě při příští prezentaci stejného podnětu zlepšil, tj. aby se rozdíl mezi aktuálním a požadovaným výstupem zmenšil. Tato procedura se provádí se všemi vstupně-výstupními dvojicemi, které chceme, aby se neuronová síť naučila. Celý proces se pak mnohokrát opakuje. V ideálním případě si takto neuronová síť osvojí

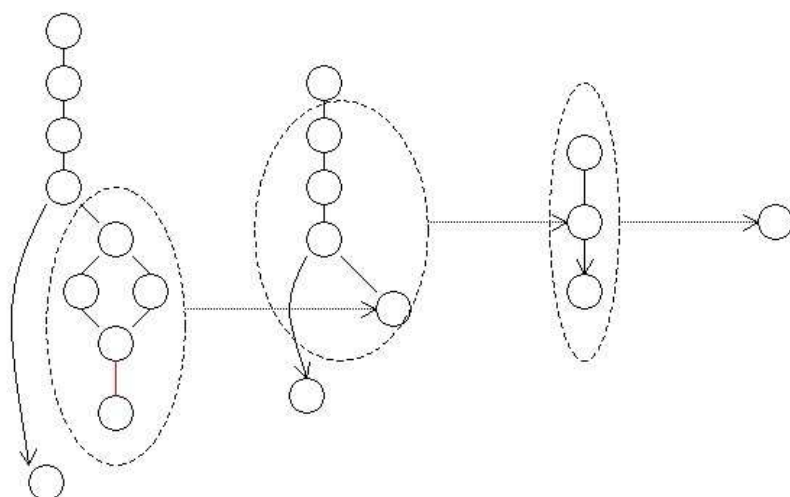
¹ Kromě změn efektivit synaptických spojů dochází v lidském mozku v rámci učení také k tzv. *synaptickému remodelingu*, kdy nervové buňky pod vlivem vnitřních a zevních podnětů budují nové synapse, nebo naopak existující synapse ničí (Koukolík, 2005).

soubor vah spojuj mezi jednotlivými uzly sítě, který jí umožní provádět požadované transformace (mapování) vstupů na výstupy.

2.2 Konekcionistické modely

Každý systém, ať už přirozený nebo umělý, se skládá z určité množiny komponent: z různých prvků, z jejich různých vlastností a z jejich vzájemných vztahů. Při snaze porozumět nějakému systému si člověk většinou pomáhá vytvořením jeho zjednodušeného modelu, do kterého jsou zahrnuty pouze ty komponenty modelovaného systému, které člověk považuje za důležité z hlediska porozumění předmětu jeho zájmu. Člověk při tom vždy stojí před problémem výběru takových komponent modelovaného systému, které na jedné straně budou dostatečně komplexní a reprezentativní, aby člověku umožnily pochopit fungování modelovaného systému, a na straně druhé budou dostatečně jednoduché, aby mu umožnily smysluplnou analýzu modelovaného systému, který se jinak skládá z milionů různých komponent. Takový dobrý model je tak vždy kompromisem mezi jeho reprezentativností a srozumitelností.

Proces modelování je významně ztížen tím, že každý systém lze vždy analyzovat na mnoha různých úrovních abstrakce. Například člověka lze analyzovat na úrovni orgánových soustav, jednotlivých orgánů, tkání, buněk, molekul, atomů, nukleonů a elektronů nebo kvarků. Tento postupný regres jednotlivých subsystémů existujících na různých rovinách abstrakce je schematicky zachycen na obrázku 61: Skupina komponent je vždy ve směru zleva doprava vnímána jako jeden nerozborný celek, který je ohraničen něčím na způsob buněčné membrány. Podle kontextu se člověk může vnitřní strukturou daného celku buďto zabývat a brát na ni zřetel, nebo ji naopak ignorovat. Většina lidí takto souhlasí s tím, že všechno, co dělá člověka člověkem, lze v principu popsat v termínech orgánů, buněk, molekul, atomů..., a to přesto, že nikdo tímto způsobem o sobě nepřemýšlí, neboť ke svému běžnému „fungování“ člověk tento druh „sebeporozumění“ nepotřebuje: pro jeho každodenní „provoz“ jsou tyto úrovně popisu a porozumění irelevantní a bylo by jen zbytečným plýtváním času se jimi zabývat. Je tomu jako v případě zákazníka aerolinek: Pasažér v letadle obvykle nechce nic vědět o tom, kolik je paliva v nádržích, jaká je rychlost větru, kolik obědů letušky během letu naservírovaly nebo jak hustý je letecký provoz ve vzdušném prostoru nad letištěm – tohle všechno pasažér nechává na jednotlivých zaměstnancích z různých úrovní hierarchické struktury aerolinek a sám se zajímá pouze o to, že se chce jednoduše dostat z bodu A do bodu B. Jakmile se ale něco v chodu aerolinky zadrhne – například se někde



Obrázek 61: *Různé roviny abstrakce při studiu přirozených nebo umělých systémů. (s mírnými úpravami převzato z Hofstadter, 2000)*

ztratí pasažérova zavazadla –, pasažér si okamžitě začne být vědom spletné hierarchie jednotlivých úrovní, které

umožňují jeho jednoduché a prosté přesouvání se z bodu A do bodu B. Podobně když onemocníme a jdeme si k lékaři pro pomoc a pro radu, je lékař nucen se na nás podívat z hlediska nižších úrovní abstrakce, než na jakých o sobě běžně uvažujeme a přemýšlíme.

Proces modelování nějakého systému tak kromě rozhodování o tom, jaké prvky a jaké vztahy mezi nimi budou zahrnuty do zjednodušeného modelu, obsahuje také rozhodování o tom, na které rovině abstrakce se budou tyto prvky a vztahy mezi nimi vůbec hledat. Pět principů zpracovávání informací v lidské mozku – uvedených v předchozím oddíle – je výsledkem právě takovýchto rozhodnutí a umělé neuronové sítě představují zjednodušený matematický model těchto teoretických předpokladů týkajících se výpočetní činnosti lidského mozku.

2.2.1 Matematické modelování dynamických systémů

Konekcionistické modelování kognitivní procesů využívá strategie matematického modelování dynamických systémů - tj. systémů, které se v určitém časovém rámci nějakým významným způsobem mění a vyvíjí. Příkladem takového dynamického systému může být počasí, sluneční soustava, srdce, celulární automat, konečný automat, Turingův stroj nebo umělá neuronová síť - ve skutečnosti je obtížné najít něco, co by nebylo dynamickým systémem. Matematické modelování dynamického systému spočívá v popisu stavů systému prostřednictvím řady numerických proměnných a ve formulaci matematických, (například diferenciálních) rovnic zachycujících dynamiku chování a vzájemnou interakci těchto proměnných v čase.

2.2.1.1 Teorie dynamických systémů

Vyjádřeno ve formální notaci, dynamický systém lze modelovat jako

$$\langle T, S, \phi_t \rangle$$

tj. prostřednictvím uspořádané časové množiny $T = \{t_0, t_1, t_2, \dots, t_n\}$, stavového prostoru $S = \{x_0, x_1, x_2, \dots, x_n\}$, a vývojového operátoru $\phi: S \rightarrow S$, který transformuje počáteční stav $x_0 \in S$ v čase $t_0 \in T$ na jiný stav $x_t \in S$ v čase $t \in T$ (Beer, 2000). Časová množina T může být diskrétní, nebo kontinuální; stavový prostor S může být numerický, nebo symbolický, kontinuální, nebo diskrétní a v závislosti na počtu proměnných nutných k popisu chování systému může mít konečný, nebo nekonečný počet rozměrů (Van Gelder, Port 1995; Beer, 2000). Z matematického hlediska se dynamický systém skládá ze dvou hlavních částí (Abraham, 1995; Franklin, 2001): Tou první je tzv. **globální dynamika systému**, což je soubor funkcí, pravidel či rovnic (F), které specifikují změnu systémových proměnných v čase. Z tohoto hlediska je dynamickým systémem jakýkoli dostatečně dobře specifikovaný soubor pravidel. Clayton (1997) jako příklad takového souboru pravidel uvádí dvojici rovnic

$$\begin{aligned}x_{n+1} &= x_n + y_n \\ y_{n+1} &= x_n\end{aligned}$$

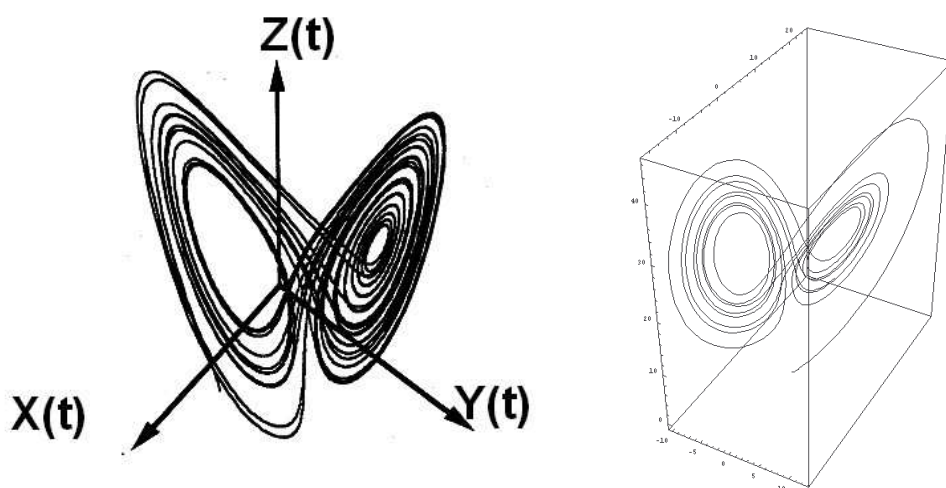
které definují jednoduchý dynamický systém se dvěma proměnnými, kde se proměnná x mění tak, že bere svou starou hodnotu a přičítá ji k hodnotě proměnné y a y se mění tak, že se stává starou hodnotou x . Globální dynamika (tj. soubor pravidel či rovnic) transformuje aktuální stav dynamického systému: $x_{n+1} = F(x_n)$ a tím mění „polohu“ systému v jeho tzv. **stavovém (fázovém) prostoru**, který představuje druhou podstatnou složkou dynamického systému.

Stavový prostor je množinou všech možných stavů, ve kterých se systém jako celek může kdy nacházet. Jednotlivé prvky, vlastnosti či proměnné systému jsou popsány jako reálná čísla, takže stav systému například se šesti proměnnými může být zachycen pomocí řady šesti čísel, čili **vektorem** o délce 6. Z tohoto hlediska tedy dynamický systém představuje systém s numerickými stavy, které se vyvíjejí v čase podle určitého pravidla (F). Stavový prostor takového systému je n -vektorovým prostorem, tj. prostorem konečných posloupností n čísel, která definují hodnotu systémových proměnných; stav systému je pak dán n -vektorem, kde n je počet systémových proměnných. Tento n -vektorový prostor lze chápat rovněž jako n -dimenzionální prostor, kde n je opět počet systémových proměnných, jejichž konkrétní hodnoty specifikují (stejně jako souřadnice na mapě) pozici stavu systému v n -dimenzionálním prostoru, jehož každý bod odpovídá jednomu možnému stavu systému.

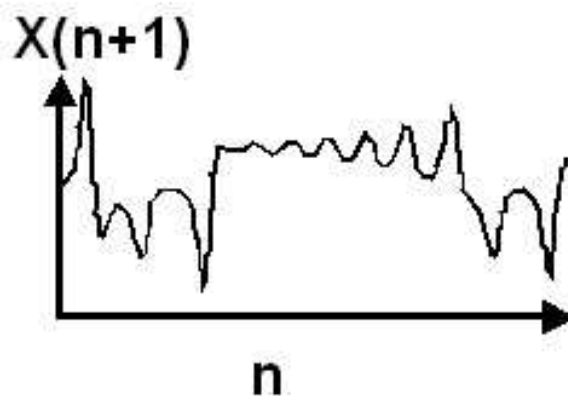
Aplikace globální dynamiky (F) na aktuální stav systému vede ke změně stavu systému, tj. ke změně polohy v n -dimenzionálním stavovém prostoru. Tento proces výpočtu nového stavu *diskrétního dynamického systému* (jehož stav se mění v diskrétních časových krocích, na rozdíl od kontinuálního dynamického systému) se nazývá **iterace**: Máme daný počáteční stav x_0 a jak čas ubíhá, globální dynamika (F) generuje posloupnost stavů:

$$\begin{aligned} F(x_0) &= x_1 \\ F(x_1) &= x_2 \\ F(x_2) &= x_3 \\ F(x_3) &= x_4 \dots \end{aligned}$$

kterou lze - v případě sledování změny vztahů mezi několika systémovými proměnnými najednou (viz obrázek 62) - zobrazit pomocí **fázového diagramu** nebo **časového rozvoje** (nazývaného také **časová řada**) - to v případě sledování změny hodnot jedné systémové proměnné (viz obrázek 63).



Obrázek 62: Fázový diagram zobrazující pohyb systému ve stavovém prostoru definovaném třemi systémovými proměnnými (X , Y , Z): V každém časovém okamžiku je pozice systému v třírozměrném prostoru specifikována aktuálními hodnotami tří proměnných (X , Y , Z); jak se systém mění a vyvíjí, vykresluje v prostoru specifickou trajektorii, která je stopou historie změn, kterými prochází hodnoty tří systémových proměnných v čase. Zde je zachycen známý Lorenzův „podivný atraktor“, který je grafickým vyjádřením chování jednoduchého modelu zvláštního druhu proudění plynů a kapalin (tzv. konvekce, kdy horké plyny a kapaliny mají tendenci stoupat směrem vzhůru). Lorenz tento typ proudění plynů a kapalin popsal pomocí sady tří diferenciálních rovnic: $\frac{dX}{dt} = \sigma(Y - X)$, $\frac{dY}{dt} = -XZ + rX - Y$, $\frac{dZ}{dt} = XY - bZ$. Výsledkem počítání systému těchto tří rovnic (s hodnotami parametrů $r = 28$, $\sigma = 10$, $b = 8/3$) je výše zmiňovaný Lorenzův „podivný atraktor“, který sleduje „podivný, výrazný tvar, jakousi dvojitou spirálu v trojrozměrném prostoru, připomínající motýlí křídla [a signalizující] naprostou neuspořádanost, neboť ani jeden bod obrazce se neopakuje, [zároveň ovšem signalizující] nový druh uspořádanosti.“ (Gleick, 1996, s. 35).



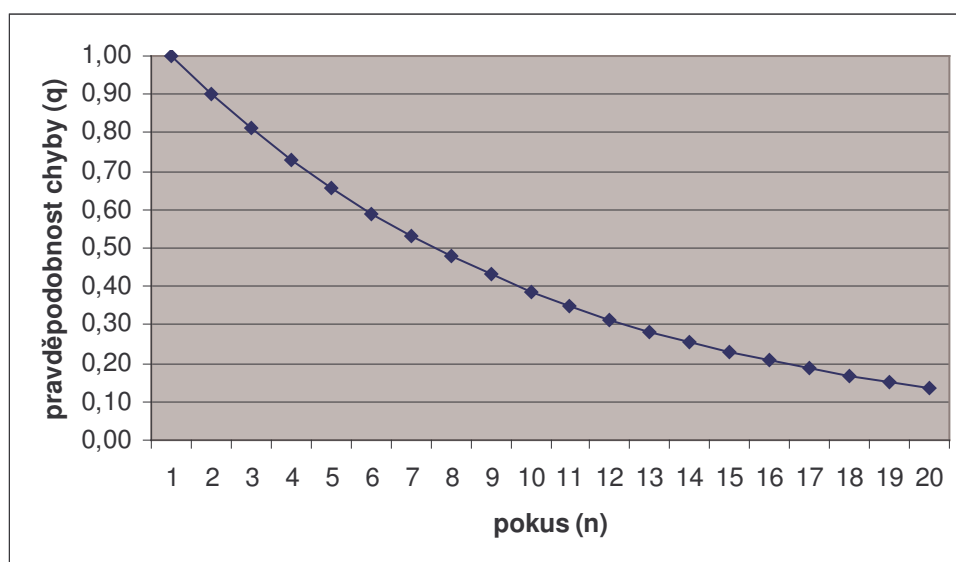
Obrázek 63: Časový rozvoj proměnné X z Lorenzova modelu konvekce.

Posloupnost změn stavu systému vytváří ve stavovém prostoru určitou **trajektorii**, **itinerář** či **orbitu dlouhodobého chování systému v čase**, což je právě to, co nás na dynamických systémech nejvíce zajímá: *Jak se bude posloupnost stavů vyvíjet jak půjdeme dál v čase? Stabilizuje se v nějakém pevném bodě? Bude donekonečna opakovat nějaký malý vzorec stavů?* Abychom mohli na takové otázky odpovědět, potřebujeme znát *globální dynamiku systému*, hodnoty tzv. *řídících parametrů*, které jsou v průběhu vývoje systému stabilní a nijak se nemění, a nakonec *počáteční podmínky* či *počáteční stav* systému, který je specifikován aktuálními hodnotami jeho proměnných x, y, z, \dots . Takto např. můžeme modelovat dynamiku chování paměťového systému pomocí jednoduchého matematického modelu (Clayton, 1997), který specifikuje, jak se v čase od jednoho pokusu k druhému mění pravděpodobnost chybné reprodukce (q) při daném pokusu (n): Globální dynamika takového paměťového systému by mohla být dána rovnicí $q(n+1) = \beta q(n)$; při počáteční stavu $q(1) = 1$ a parametru učení $\beta = 0.9$ by pak aplikace globální dynamiky na výchozí stav paměťového systému vedla k následující posloupnosti stavů...

$$\begin{aligned} q(1) &= 1 \\ q(2) &= \beta q(1) = (0.9)(1) = 0.9 \\ q(3) &= (0.9)q(2) = (0.9)(0.9) = 0.81 \\ &\dots \end{aligned}$$

Celkové chování takto specifikovaného paměťového systému v čase je zachyceno v grafu 1.

Existují tři základní druhy dlouhodobého chování dynamických systémů v čase – **pevný bod**, **periodické chování** a **chaotické chování**. Franklin (2001, s. 281) tyto možné druhy dlouhodobého chování dynamických systémů ilustruje na nekonečných posloupnostech reálných čísel mezi 0 a 1: Posloupnosti jako 0.3333333333333333..., 0.1233333333333333... nebo 0.9876543211111111... směřují k jednomu **pevnému bodu**



Graf 1: Časový rozvoj pravděpodobnosti chybné reprodukce pro model paměťového výkonu za předpokladu $q(1)=1, \beta=0.9$. (převzato z Clayton, 1997)

(fixed point); posloupnosti 0.123123123123123... či 0.6845901212121212... vykazují **periodické chování s periodou n** ; a konečně chování posloupností 0.41421356237309... nebo 0.14159678763244... je **chaotické a nepředpověditelné**, přesto plně deterministické - každý bod v orbitě lze vypočítat aplikací globální dynamiky (T) na předchozí stav; chaotická orbita je nepředpověditelná pouze v tom smyslu, že neexistuje žádný rychlejší způsob, jak určit její chování, než spočítat jeden její stav za druhým (tzv. *deterministický chaos*).

V souvislosti s dlouhodobým chováním dynamických systémů se také někdy hovoří o tzv. **limitních množinách - bodech či cyklech** - dynamického systému, které představují ty stavy systému, ke kterým má systém tendenci směřovat. Franklin (2001, s. 282) tento koncept ilustruje na příkladě dynamiky jednoduchého dynamického systému, jehož stavový prostor tvoří všechna reálná čísla a jehož globální dynamika $F(x) = x^2$. Níže je uveden přehled trajektorií ve stavovém prostoru, které mohou generovat různé počáteční stavy takového systému...

$$X_0 = 0 \rightarrow 0$$

$$X_0 = 1 \rightarrow 1$$

$$X_0 > 1 \rightarrow \dots \infty$$

$$X_0 = (0, 1) \rightarrow \dots 0$$

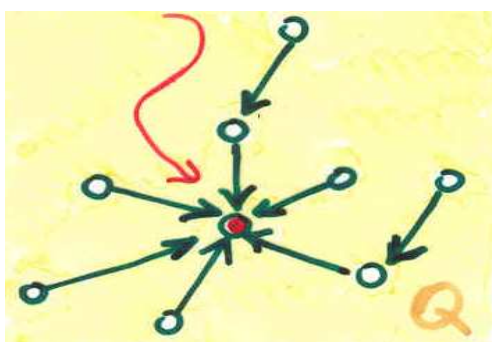
$$X_0 = (-1, 0) \rightarrow \dots 0$$

$$X_0 = -1 \rightarrow 1$$

$$X_0 < -1 \rightarrow \dots \infty$$

Z tohoto výčtu možných trajektorií je jasné patrné, že limitními body tohoto dynamického systému, ke kterým mají jeho vývojové trajektorie tendenci směřovat, jsou 0, 1 a ∞ .

Tyto limitní body dynamického systému jsou známější pod názvem **atraktory**, což jsou body či oblasti ve stavovém prostoru, které přitahují okolo procházející trajektorie (viz obrázek 64). **Bodovými atraktory** výše uvedeného systému jsou 0 a ∞ - to jsou body ve stavovém prostoru, které si k sobě stahují většinu možných trajektorií. Abychom získali nějaký **cyklický atraktor**, museli bychom mírně pozměnit globální dynamiku výše

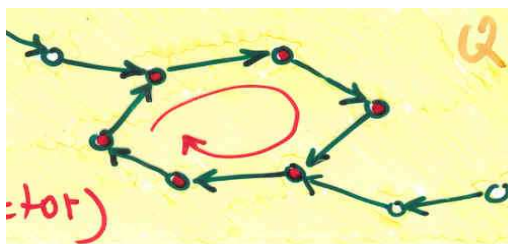


Obrázek 64: Schématické znázornění atraktoru dynamického systému. (s mírnými úpravami převzato z ppt prezentace Ivana M. Havla k přednáškám „Přirozené a umělé myšlení“ konaným ve školním roce 2005/2006 na pracovišti Centra Teoretických Studií v Praze)

uvedeného systému tak, aby $F(x) = x^2 - 1$. Trajektorie takového systému by pak vypadaly následujícím způsobem...

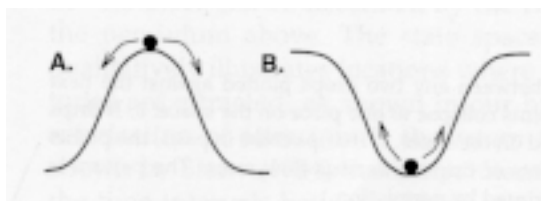
$$\begin{aligned} X_0 = 0 &\rightarrow -1, 0, -1, 0, -1, 0, \dots \\ X_0 = 1 &\rightarrow 0, -1, 0, -1, 0, -1, \dots \\ X_0 > 1 &\rightarrow \dots \infty \\ X_0 = (0, 1) &\rightarrow \dots -1, 0, -1, 0, -1, \dots \\ X_0 = (-1, 0) &\rightarrow \dots -1, 0, -1, 0, -1, \dots \\ X_0 = -1 &\rightarrow 0, -1, 0, -1, 0, -1, \dots \\ X_0 < -1 &\rightarrow \dots \infty \end{aligned}$$

Takto definovaný dynamický systém má tendenci směřovat především k cyklickému atraktoru $-1/0$, tedy ke smyčce s periodou 2, kdy se systém pravidelně navrácí ve dvou krocích do stejného stavu.



Obrázek 65: Schématické znázornění cyklického atraktoru dynamického systému. (s mírnými úpravami převzato z ppt prezentace Ivana M. Havla k přednáškám „Přirozené a umělé myšlení“ konaným ve školním roce 2005/2006 na pracovišti Centra Teoretických Studií v Praze)

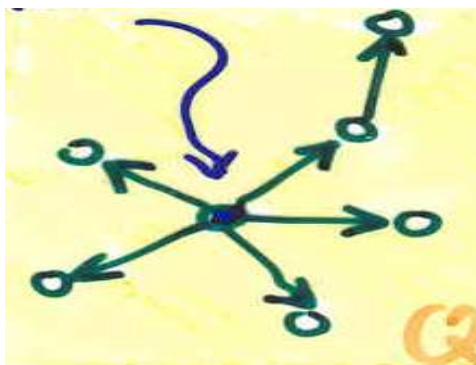
Každý atraktor definuje svou **basin of attraction**¹, což je oblast ve stavovém prostoru, která spadá pod „sféru vlivu“ daného atraktoru, takže kdykoli se systém v této oblasti ocitne, vždy se nakonec usadí v limitním bodu či cyklu definovaném daným atraktorem. Basin of attraction si lze představit nejlépe jako určitým způsobem zakřivený prostor, kde dno údolí představuje atraktor a svahy údolí jeho basin of attraction (viz obrázek 66); kulička, která představuje aktuální stav systému, se pak v tomto prostoru bude pohybovat v závislosti na jeho topologii – takže když se kulička bude nacházet např. na svahu nějakého údolí, bude mít tendenci se přes posloupnost různých pozic skutálet na dno údolí. V původním numerickém systému (s globální dynamikou $F(x) = x^2$) jsou basin of attraction atraktoru 0 všechna čísla mezi 1 a -1 a v případě atraktoru ∞ všechna čísla větší než 1 a všechna čísla menší než -1.



Obrázek 66: Chování dynamického systému v čase, resp. jeho „pohyb“ ve stavovém prostoru si lze názorně představit jako pohyb kuličky, jejíž chování je determinováno zakřivenou topologií prostoru, v němž se pohybuje.

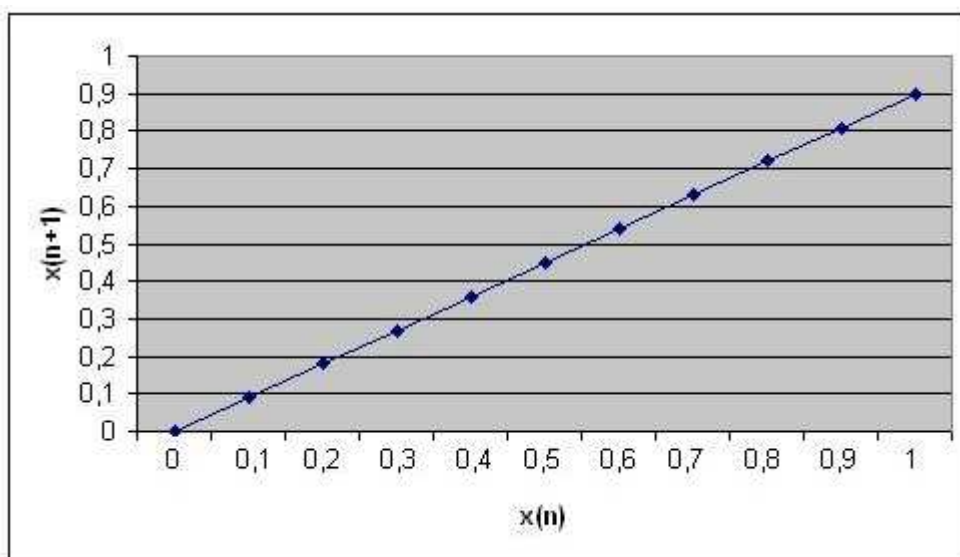
Vedle atraktorů existují ve stavovém prostoru také tzv. **repelory** (doslova *odpuzovače*), což jsou body či oblasti ve stavovém prostoru, které okolo procházející trajektorie – na rozdíl od atraktorů – odklání. V případě původního jednoduchého dynamického systému (s globální dynamikou $F(x) = x^2$) je takovým repelorem 1 - kdykoli se octneme v její blízkosti, skončíme vždy buď u 0 nebo u ∞ .

¹ Doslova miska, pánev či kotlina přitažlivosti, což je překlad, který na jednu stranu velice dobře vystihuje podstatu tohoto konceptu (viz hlavní text), na druhou stranu působí dosti neumělým dojmem (na „třetí stranu“ nepůsobí o nic více neuměle než „černá díra“ nebo „kvark“, takže celé je to možná pouze otázka zvyku); proto - a navíc také z důvodu neexistence nějakého ustáleného českého překladu - budu v textu při těch několika málo příležitostech i nadále používat původní (nesklonný) anglický termín.



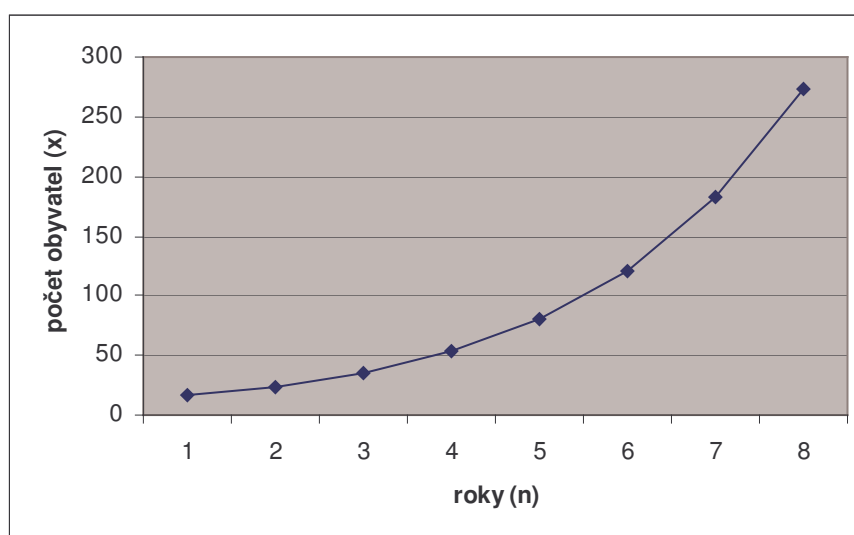
Obrázek 67: Schématické znázornění repeloru dynamického systému. (s mírnými úpravami převzato z ppt prezentace Ivana M. Havla k přednáškám „Přirozené a umělé myšlení“ konaným ve školním roce 2005/2006 na pracovišti Centra Teoretických Studií v Praze)

Důležitou vlastností většiny zajímavých dynamických systémů je jejich **nelinearita**. Globální dynamiku či funkci specifikující chování takového nelineárního dynamického systému není možné v grafu zakreslit jako přímku, neboť hodnota výstupu funkce zde není úměrná hodnotě vstupu. To, co přitom činí dynamický systém nelineárním, není nelinearita jeho chování, ale nelinearita funkce, která jeho chování specifikuje (Clayton, 1997). Tak se například výše zmiňovaný paměťový model může na první pohled jevit jako nelineární dynamický systém, neboť graf zachycující jeho chování nemá podobu přímky (viz graf 1). Ovšem tento graf zachycuje chování modelovaného systému, nikoli funkci $(q(n+1) = \beta q(n))$, která chování systému determinuje. Když vytvoříme graf funkce tohoto systému $(q(n+1))$ vs. $q(n)$, uvidíme, že výsledná křivka má podobu přímky, tedy že $q(n+1)$ je lineární funkcí $q(n)$ (viz graf 2).

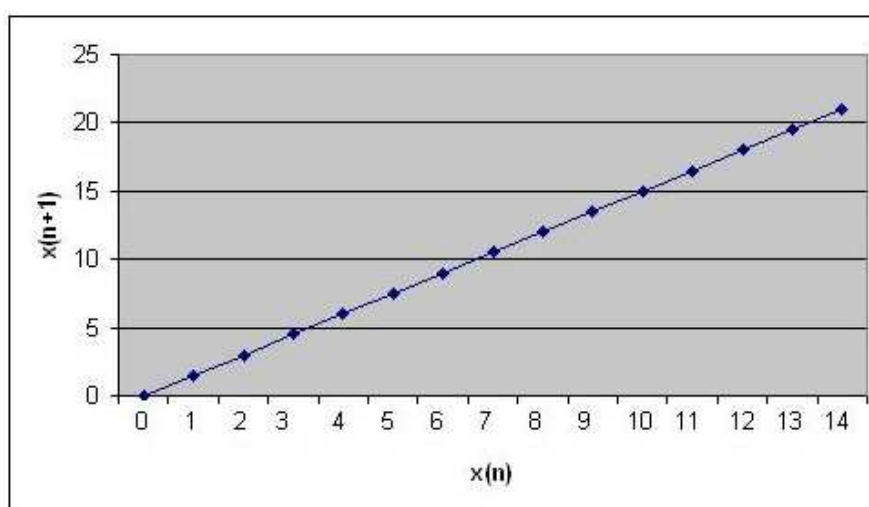


Graf 2: Paměťová funkce, $x(n+1)$ vs. $x(n)$, $\beta = 0,9$. (převzato z Clayton, 1997)

Chování takového nelineárního dynamického systému se obvykle demonstruje na příkladu tzv. *logistické diferenciální rovnice*, které se také někdy říká *logistická mapa* (Clayton, 1997): Logistická mapa je mírně upravenou verzí jednoduchého modelu růstu, jehož globální dynamika je definována rovnicí $x(n+1) = r x(n)$; x se mění od jednoho časového okamžiku (n) k druhému ($n+1$) v závislosti na parametru r ; když $r > 1$, x se s postupnou iterací modelu neustále zvětšuje; když $r < 1$, x se bude neustále zmenšovat. V grafu 3 je zachyceno chování modelovaného systému pro $r = 1,5$. Populace v roce 2 čítá 16 obyvatel ($x(1)=16$); a protože růstový parametr $r = 1,5$, populace (x) se každý rok zvětšuje o 50 %. Jinými slovy, populace exponenciálně narůstá - již v 25. roce by přesáhla hranici čtvrt milionu obyvatel. Tento jednoduchý model růstu je lineární (viz graf 4) a jeho výsledkem je neomezený růst (nebo pokles, to podle velikosti řídicího parametru). Abychom získali nelineární model růstu,

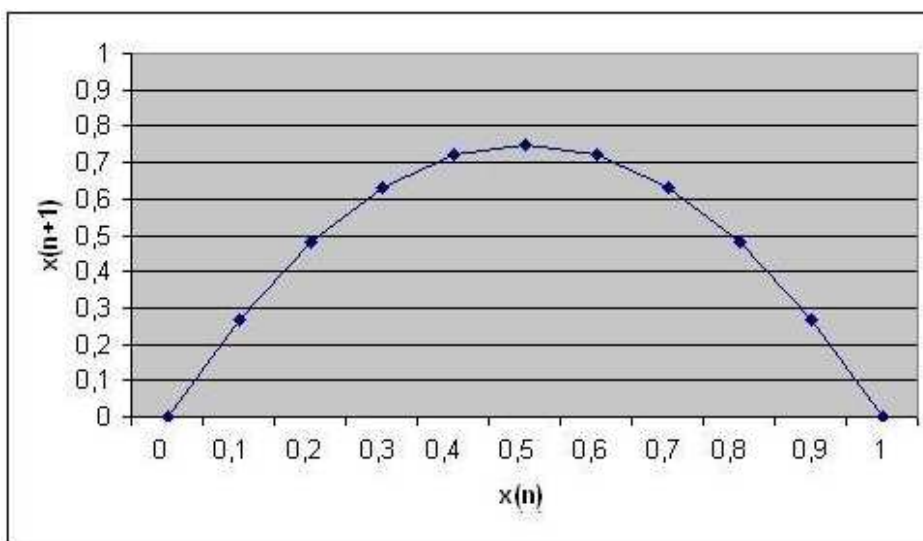


Graf 3: Iterace jednoduchého modelu růstu, $r = 1,5$. (převzato z Clayton, 1997)

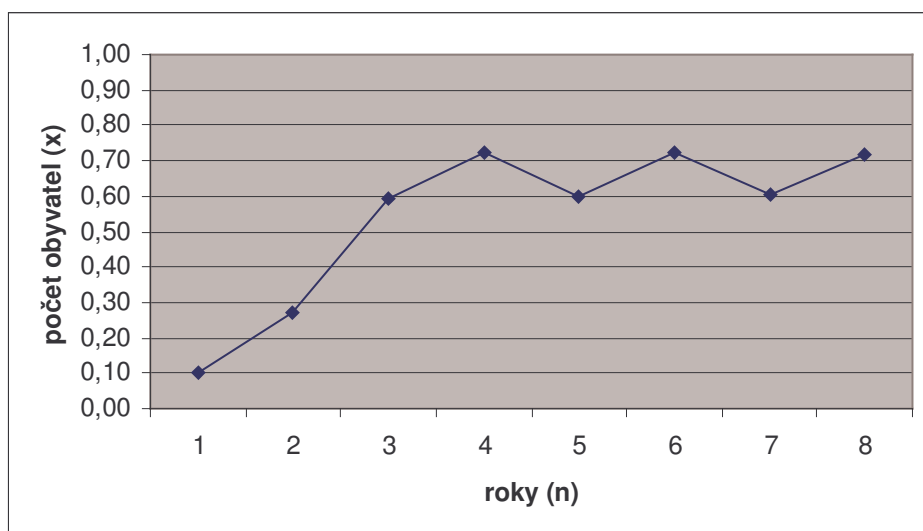


Graf 4: Lineární funkce jednoduchého modelu růstu, $x(n+1)$ vs. $x(n)$, $r = 1,5$. (převzato z Clayton, 1997)

který by zabránil ničím neomezenému růstu, musíme k rovnici původního modelu přidat něco, co by dokázalo růst potlačit, kdykoli by dosáhl příliš vysoké úrovně. A právě to dělá logistická mapa, která do rovnice původního modelu neomezeného růstu přidává nový člen $[1 - x(n)]$, který má tuto omezovací funkci. Globální dynamika logistické mapy vypadá následovně: $x(n+1) = r x(n) [1 - x(n)]$. Nový člen $[1-x(n)]$ v ní zabraňuje neomezenému růstu tím, že jak se hodnota x blíží 1, $[1-x(n)]$ se blíží 0. Grafické znázornění této nelineární funkce pro $r = 3$ je znázorněno v grafu 5. Vývoj populace řídící se globální dynamikou logistické mapy pro $r = 3$ a $x(1) = 0,1$ je pak zachycen v grafu 6.

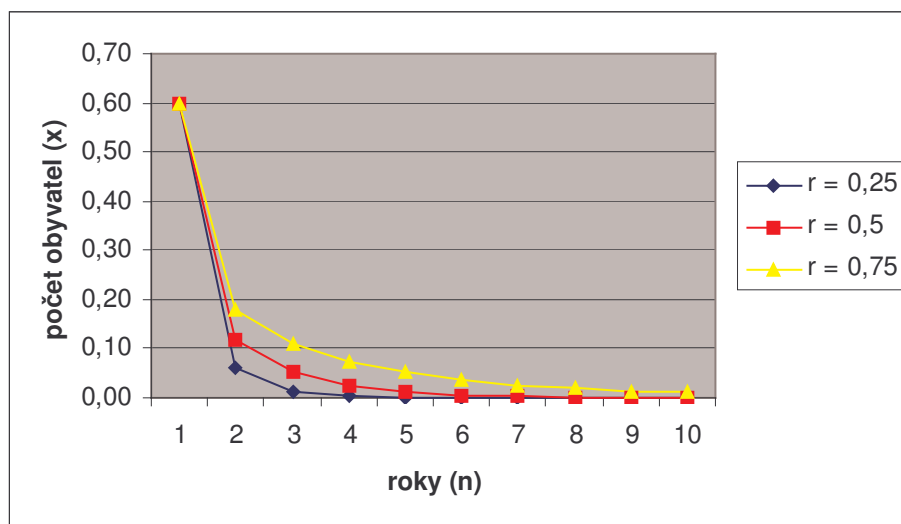


Graf 5: Logistická mapa, $x(n+1)$ vs. $x(n)$, $r = 3$. (převzato z Clayton, 1997)

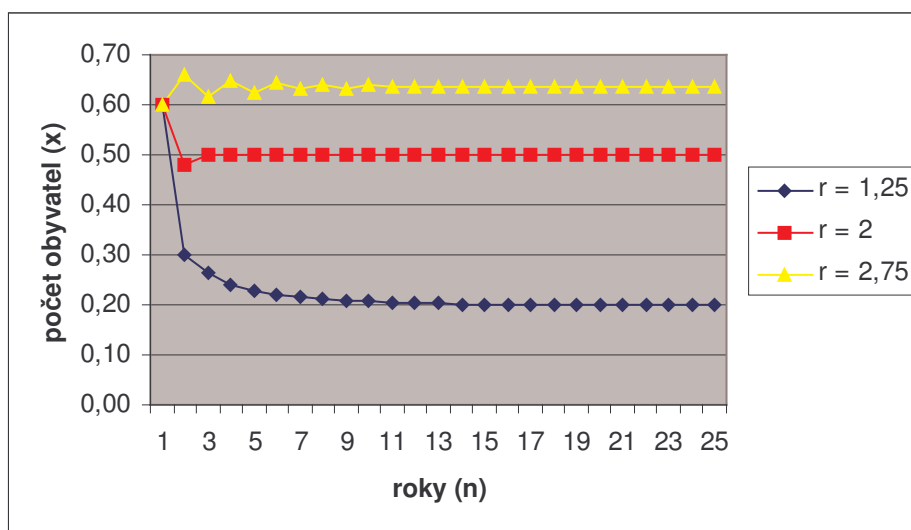


Graf 6: Chování logistické mapy pro $r = 3$, $x(1) = 0,1$. Míra růstu (x) je proporčně překalibrována tak, že maximální hodnota, které x může dosáhnout, je rovna 1. (převzato z Clayton, 1997)

Pro chování nelineárních dynamických systémů je charakteristické, že velice **citlivě reaguje na jakékoli změny hodnot klíčových řídicích parametrů** systému. Tuto citlivost lze názorně ilustrovat na způsobu, jakým se začne měnit chování logistické mapy v závislosti na změnách hodnoty jejího řídicího parametru r : Když $r < 1$, systém má tendenci se ustálit na nulových hodnotách proměnné x . V grafu 7 můžeme vidět tři různé trajektorie, které odpovídají třem různým hodnotám parametru r ; přestože se jejich individuální průběh liší, všem třem je společné směřování k nulovým hodnotám proměnné x , na kterých se po nějaké době také ustálí. Tento typ chování je příkladem **nulového jednobodového atraktoru**. Chování logistické mapy při $r = (1,3)$ je charakterizováno řadou **nenulových jednobodových atraktorů**; systém má tendenci se ustálit (v závislosti na hodnotě parametru r) v jednom konkrétním nenulovém stavu (viz graf 8).

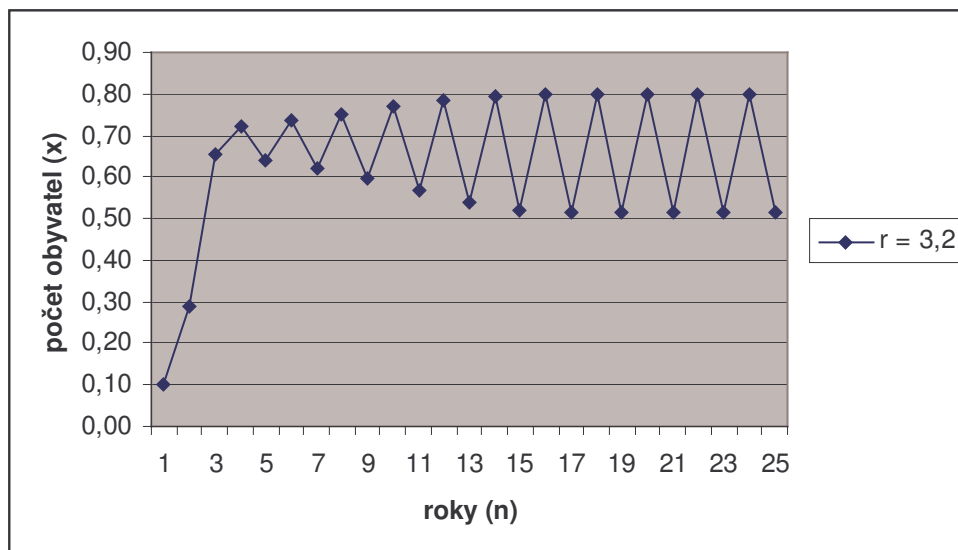


Graf 7: Chování logistické mapy pro $r=0,25, 0,50$ a $0,75$ a $x(1)=0,6$. (převzato z Clayton, 1997)

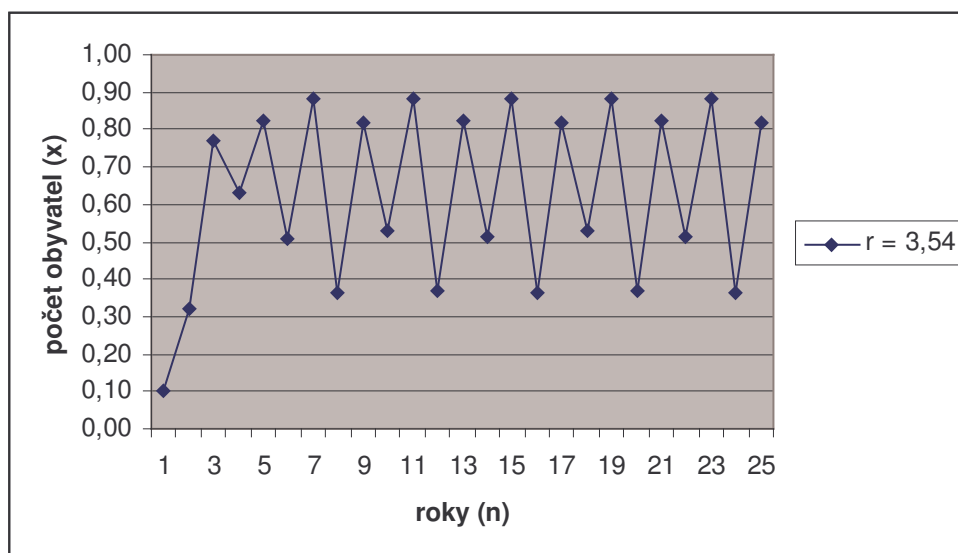


Graf 8: Chování logistické mapy pro $r=1,25, 2$ a $2,75$ a $x(1)=0,6$. (převzato z Clayton, 1997)

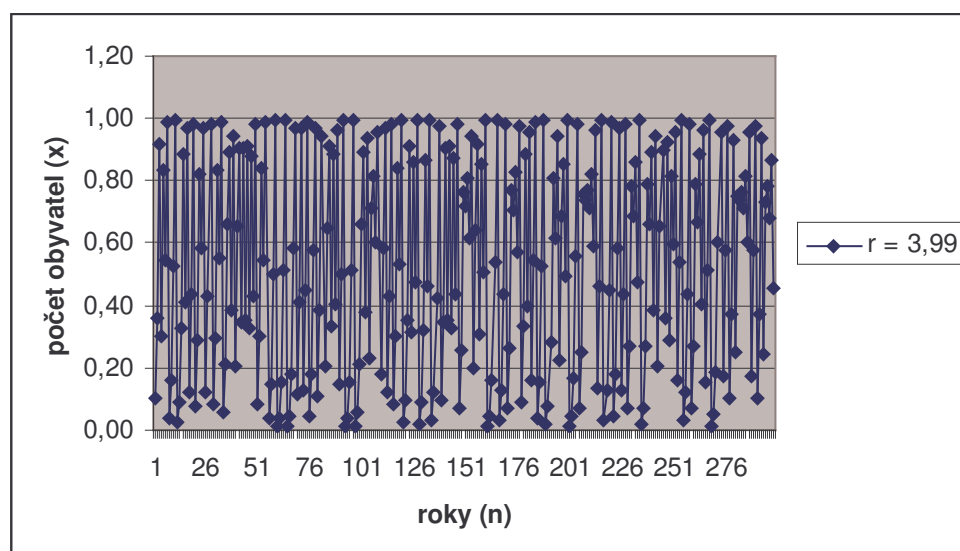
Při $r > 3$ se chování logistické mapy stává zajímavější; můžeme tu vidět **dvoubodový** či **čtyřbodový cyklický atraktor** (grafy 9 a 10) a také **N-bodový**, neboli **chaotický atraktor** (graf 11).



Graf 9: Chování logistické mapy pro $r=3,2$ a $x(1)=0,1$. (převzato z Clayton, 1997)

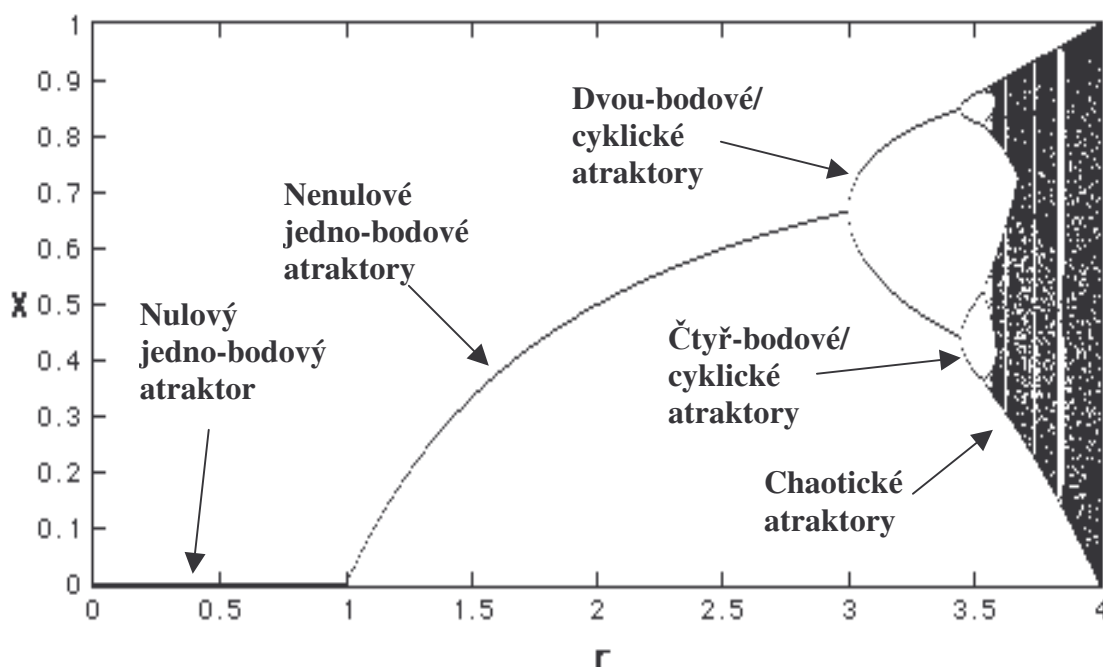


Graf 10: Chování logistické mapy pro $r=3,54$ a $x(1)=0,1$. (převzato z Clayton, 1997)



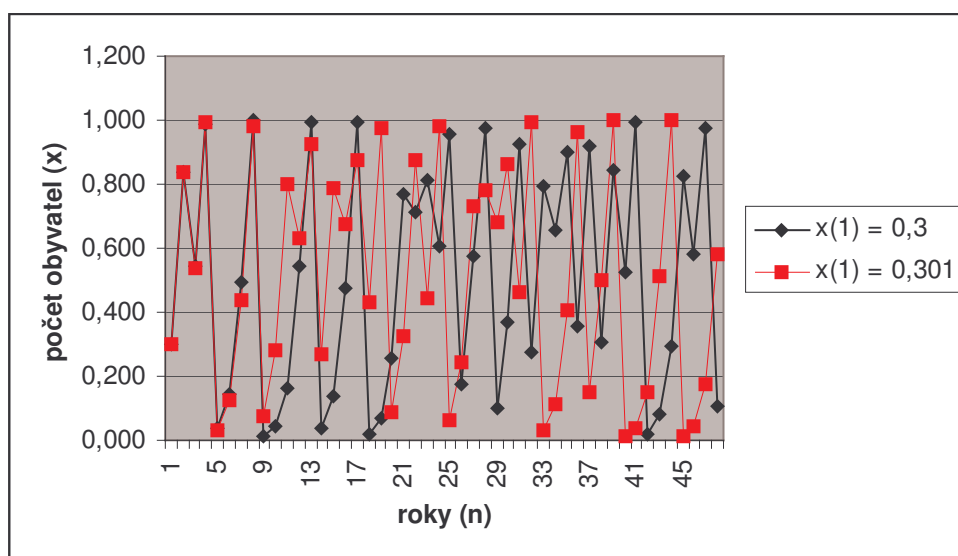
Graf 11: Chování logistické mapy pro $r=3,99$ a $x(1)=0,1$. (převzato z Clayton, 1997)

Tyto změny topologie stavového prostoru - a v důsledku toho tedy i chování systému - v závislosti na změně hodnoty řídicího parametru jsou tradičně zobrazovány pomocí **diagramu bifurkace** (viz obrázek 68), který je schematickým vyjádření přechodů (bifurkací) mezi různě „zvlněnými“ krajinami atraktorů.

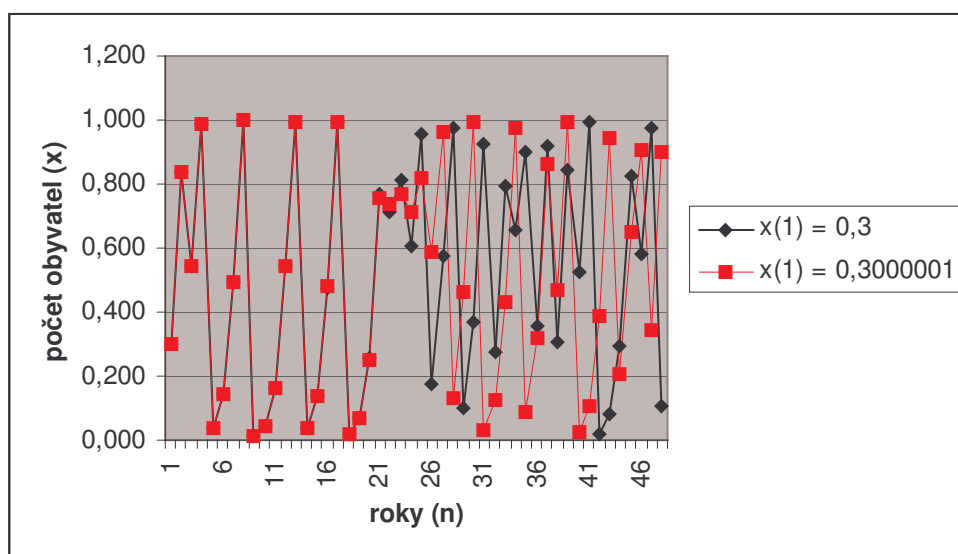


Obrázek 68: Bifurkační diagram znázorňující změnu topologie krajiny atraktorů logistické mapy v závislosti na změně řídicího parametru r .

Další důležitou vlastností nelineárních dynamických systémů je jejich **citlivost na počáteční podmínky** (tzv. *motýlí efekt*). Opět to lze demonstrovat na různém chování logistické mapy při minimálních rozdílech v počátečním stavu systému: Na grafu 12 můžeme vidět, že počáteční stavy dvou trajektorií se liší pouze v jedné tisícíně, přesto se po několika málo letech (iteracích) jejich dráhy přestávají překrývat. Dokonce i v případě, že rozdíl v počátečním stavu bude pouze jedna desetimiliontina (viz graf 13), obě trajektorie se přestanou překrývat pouze o několik málo let (iterací) později.



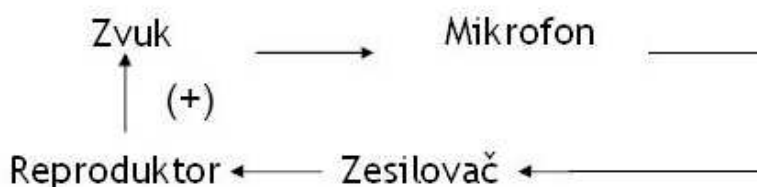
Graf 12: Časová řada logistické mapy pro $r=3,99$, $x(1)=0,3$ vs. $x(1)=0,301$ a 48 iterací. (převzato z Clayton, 1997)



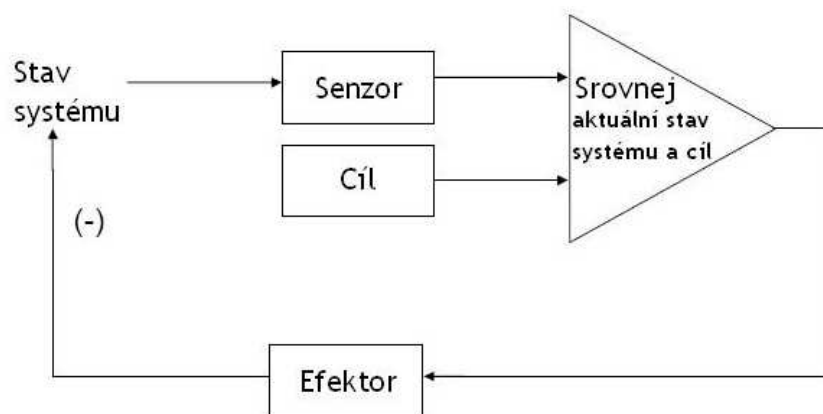
Graf 13: Časová řada logistické mapy pro $r=3,99$, $x(1)=0,3$ vs. $x(1)=0,3000001$ a 48 iterací. (převzato z Clayton, 1997)

K pochopení chování dynamických systémů jsou rovněž důležité koncepty **zpětné vazby** (feedback) a **sebeorganizace** (self-organization) či **sebeřízení** (self-control). Pro většinu dynamických systémů je charakteristické, že se jedná o sebeorganizující systémy, tj. systémy, jejichž chování není formováno a determinováno pouze podněty přicházejícími zpoza hranic systému, ale také samotným systémem. Kostroň (1997, s. 166) sebeorganizaci a sebeřízení definuje jako „stav, kdy řídicí parametr dynamického systému závisí na (je funkcí) stavu systému..., [tzn. že] jde o zpětnou vazbu mezi systémem a jeho řídicími elementy“.

Tato sebeorganizace dynamických systémů je zapříčiněna tím, že komplexní systémy jsou vždy tvořeny souborem většího množství různých vzájemně propojených a interagujících zpětnovazebních smyček, které existují mezi jednotlivými prvky systému. Existují přitom dva základní typy zpětné vazby: **pozitivní** (samoposilující) a **negativní** (samoregulující nebo také kompenzační). „Pozitivní cykly zesilují výstup systému a zvětšují odchylku výstupu od normálu a negativní cykly zase naopak změnám brání, zmenšují odchylku výstupu od normálu a usilují tak o udržení rovnováhy.“ (Mildeová, Vojtko, 2003, s. 37-38) Příkladem jednoduchého systému s pozitivní zpětnou vazbou je systém tvořený mikrofonem, zesilovačem a reproduktorem (viz obrázek 69), který způsobuje známé „vazbení“: Zvuk se přes mikrofon dostane k zesilovači, který ho zesílený odešle k reproduktoru, jehož výstup je vstupem do mikrofonu atd. Postupně se zvuk pozitivní zpětnou vazbou zesílí až do nesnesitelného a uši trahajícího kvílení. Na obrázku 70 je pak schematically znázorněna struktura jednoduchého řídicího systému, který ke své činnosti využívá negativní zpětnou vazbu. Příkladem takového jednoduchého řídicího systému je termostat, jehož funkcí je udržovat v místnosti stálou požadovanou teplotu: Termostat průběžně porovnává aktuální teplotu v místnosti s požadovanou teplotou; v případě, že mezi těmito dvěma hodnotami existuje rozdíl, pustí se termostat do své omezené akce, která spočívá ve vypnutí, resp. v zapnutí topení: je-li teplota v místnosti příliš vysoká, termostat vypne topení, je-li teplota naopak příliš nízká, termostat topení zapne. Kombinaci pozitivní a negativní zpětné vazby



Obrázek 69: Příklad pozitivní zpětné vazby.



Obrázek 70: Schematické znázornění systému s negativní zpětnou vazbou.

v jednom systému Mildeová a Vojtko (2003, s. 38) ilustrují na příkladu populace králíků, jejíž chování „lze popsat dvěma vzájemně svázanými zpětnovazebními smyčkami. V jedné smyčce – čím větší bude populace králíků, tím větší bude porodnost (pozitivní vazba) a čím větší bude porodnost, tím větší bude populace (pozitivní vazba). Dostaneme tak pozitivní zpětnovazebnou smyčku, která způsobuje geometrický růst populace králíků. V druhé smyčce – čím větší bude populace králíků, tím větší bude úmrtnost (pozitivní vazba) a čím větší bude úmrtnost (ať už přirozenou smrtí, nebo jako kořist predátorů), tím menší bude populace králíků (negativní vazba). Dostaneme tak negativní zpětnovazebnou smyčku, která působí na populaci králíků stabilizačně. Pokud převládne pozitivní smyčka, když odstraníme predátory,...populace králíků se bude zvětšovat a dojde k přemnožení. Pokud naopak převládne negativní zpětná vazba, znázorňující úbytek králíků díky přirozené úmrtnosti, lovem predátorů nebo chemizací a mechanizací v zemědělství, populace se bude zmenšovat až nakonec bude hrozit i vymizení králíků. Ve „zdravém“ ekosystému ovšem nastane rovnováha mezi oběma ději.“ Struktura zpětnovazebních vztahů mezi jednotlivými systémovými proměnnými tímto způsobem determinuje chování sebeorganizujícího dynamického systému v čase.

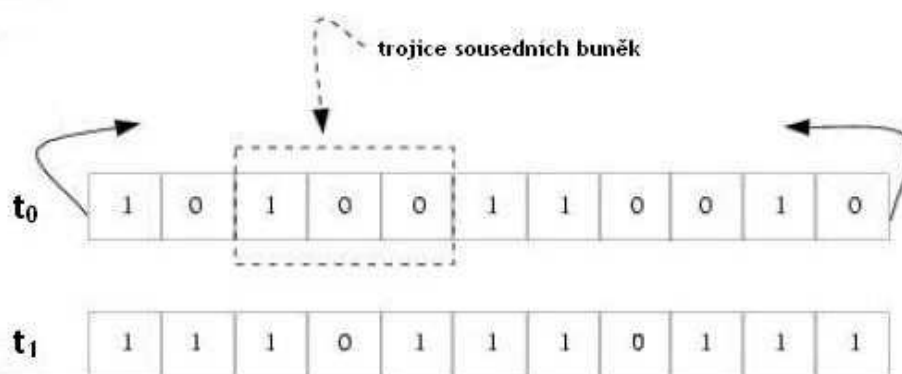
Dynamika spojená se zpracováváním informací, tj. s jejich manipulací a postupnou transformací, je dobře patrná - díky možnosti snadné vizualizace - v případě **elementárního celulárního automatu** (dále ECA) - decentralizované, distribuované a paralelní výpočetní architektury, která je realizována v pravidelné síti jednoduchých prvků s lokální konektivitou. ECA se skládá z jednodimenzionální sítě prvků - buněk -, které mohou v libovolném počtu tvořit buď otevřenou řadu, nebo uzavřenou smyčku. ECA má binární povahu, tzn. že jeho jednotlivé prvky se mohou nacházet v jednom ze dvou možných stavů: zapnuto/vypnuto, 1/0. Stavů všech buněk ECA se synchronně mění v diskrétních časových krocích na základě jejich aktuálního stavu a stavu jejich dvou nejbližších sousedních buněk. Vývojový operátor ϕ , který

mění stav buněk ECA, se nazývá *pravidlo CA* nebo *pravidlová tabulka* (Mitchell, 1998). Pro ECA existuje celkem 256 možných pravidel, která pro osm (2^3) možných vstupních trojic specifikují výstup v podobě stavu prostřední buňky. K jejich označení se zavedl jednoduchý kódovací systém, který spočívá v převodu binárního čísla daného „důsledkovou“ částí pravidla CA (čtenou ve směru od 7 (v binárním zápisu 111) k 0 (v binárním zápisu 000)) do dekadické číselné soustavy. Na obrázku 71 je takto zobrazena aplikace pravidla č. 110 (01101110) na náhodně zvolenou výchozí (stavovou) konfiguraci kruhové ECA v čase t_0 , jejímž výsledkem je změněný stav ECA v čase t_1 . Navzdory jednoduchým pravidlům, kterými se vývoj ECA v čase řídí, v důsledku své distribuované a nelineární povahy ECA produkuje složité a jen velice obtížně předpověditelné chování. Dynamiku chování ECA v delším časovém horizontu, resp. jeho stavový prostor (včetně limitních množin) lze velice snadno vizualizovat s pomocí prostoro-časového diagramu, ve kterém každý řádek odpovídá jednomu časovému kroku vývoje ECA a každý sloupec pozici jedné buňky. Na obrázku 72 je takto zachyceno chování ECA ve 200 časových krocích; ECA se skládá z 1-D sítě 200 buněk, jejichž stav se v každém časovém kroku mění podle pravidla č. 110; výchozí konfigurace

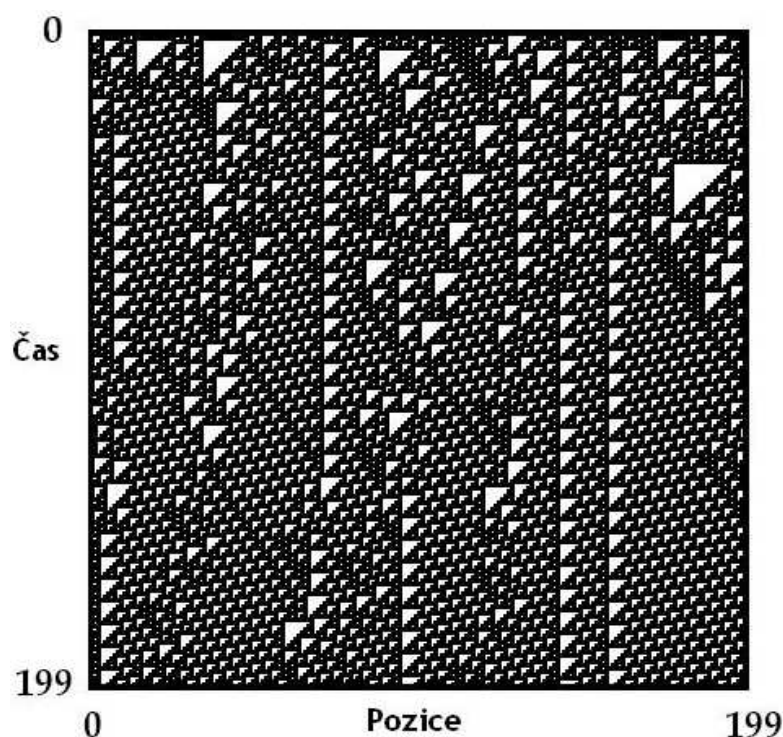
Pravidlová tabulka/vývojový operátor ϕ :

trojice sousedních buněk:	000	001	010	011	100	101	110	111
výstupní stav prostřední buňky:	0	1	1	1	0	1	1	0

Sít' buněk:



Obrázek 71: Elementární celulární automat (ECA) tvořený jednodimenzionální sítí jedenácti buněk tvořících uzavřenou smyčku. Jednotlivé buňky se mohou nacházet v jednom ze dvou možných stavů (1/0). Stav buněk ECA se mění synchronně v diskrétních časových krocích (t_0, t_1, \dots) na základě jejich aktuálního stavu a stavu dvou bezprostředně sousedících buněk podle pravidlové tabulky č. 110, která pro osm (2^3) možných vstupních trojic specifikuje výstup v podobě stavu prostřední buňky. (s úpravami převzato z Mitchell, 1998)

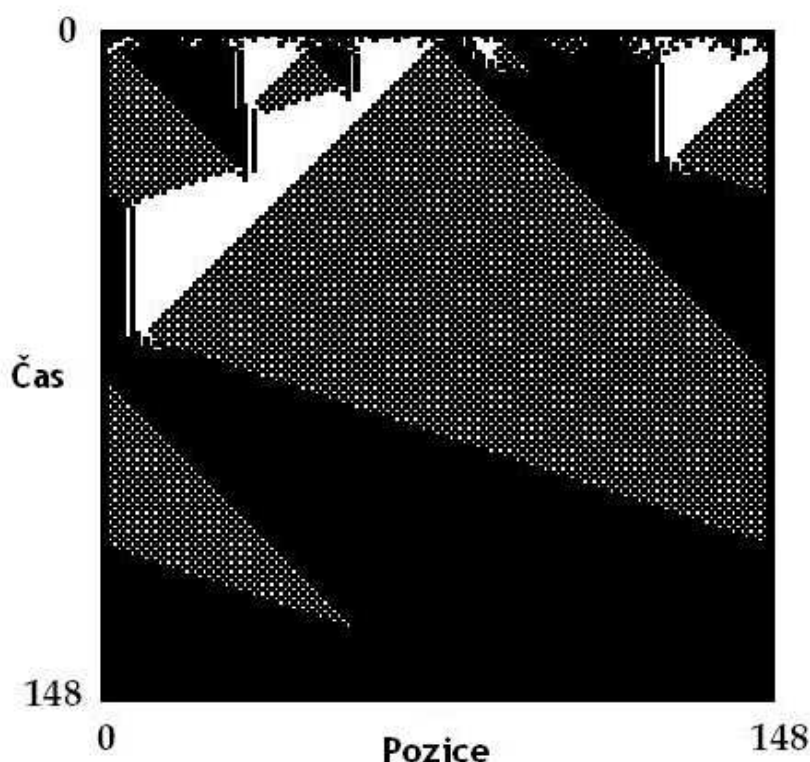


Obrázek 72: Prostorovo-časový diagram zachycující chování elementárního celulárního automatu (ECA) ve 200 časových krocích. Změna stavu buněk v jednotlivých časových krocích se mění podle pravidla č. 110 (viz obrázek 70). Dva možné stavy buněk (1/0) jsou zde reprezentovány černou, resp. bílou barvou. Každý řádek diagramu zachycuje stav ECA v jednom časové okamžiku a každý sloupec pak představuje průběh změn stavů jedné buňky v čase. Diagram jako celek pak člověku dává celkem dobrou představu o vývoji dlouhodobého chování ECA v čase. (s úpravami převzato z Mitchell, 1998)

ECA je náhodná. Pro účel ilustrace dynamického aspektu procesu zpracovávání informací je důležité, že na celulární automaty (CA) se lze dívat jako na zvláštní druh počítačů: Nazíráno z této perspektivy jsou pravidla CA počítačovým programem, výchozí konfigurace CA v čase t_0 vstupní informací, kterou je potřeba zpracovat, a vývoj stavů CA v čase je pak samotným výpočetním procesem, během něhož dochází k manipulaci a transformaci vstupní informace na požadovanou výstupní informaci, která má podobu určitého konečného vzorce stavů CA. Z mírně odlišného úhlu pohledu se lze na výchozí konfiguraci CA dívat jako na program a vstup zároveň a na pravidlo CA pak jako na hardware, na kterém běží program zpracovávající daný vstup; i zde by byl výstup výpočetního procesu kódován konečným vzorcem stavů CA. Přestože bylo teoreticky prokázáno, že některé CA¹ dokáží emulovat Turingův univerzální

¹ Celulární automaty mohou mít různou podobu danou různým počtem buněk, různým počtem dimenzí, ve kterých se tyto buňky mohou nacházet, různým počtem možných stavů nebo různými vývojovými či přechodovými pravidly. Například známá Conwayova „Hra života“ je dvojdimenzionálním binárním CA, jehož jednotlivé buňky mění svůj stav na základě svého aktuálního stavu a stavu osmi sousedních buněk (viz také s. 16-17).

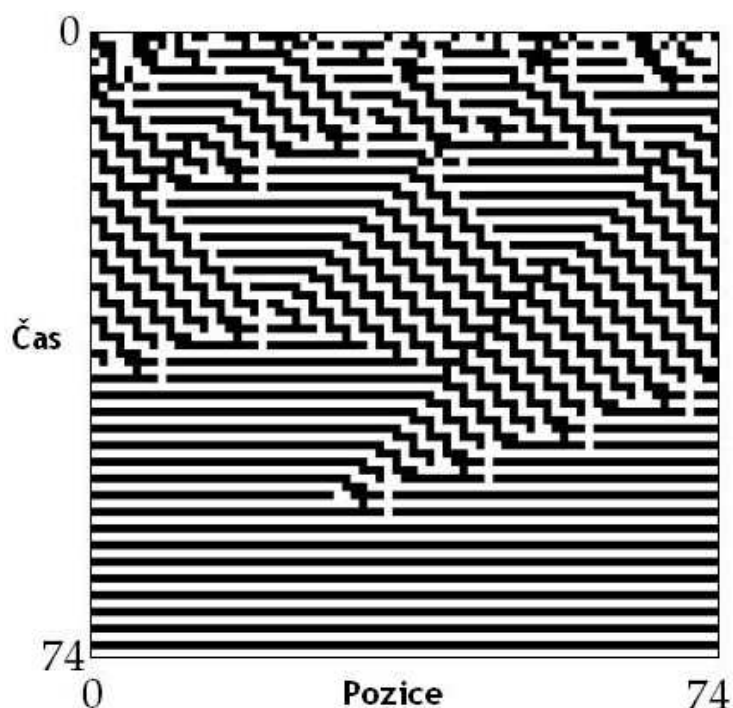
stroj, a realizovat tak tedy jakýkoli dostatečně dobře specifikovaný výpočet, prakticky je velice obtížné najít právě takové pravidlo CA a právě takovou počáteční konfiguraci, které by CA umožnily počítat nějakou zajímavou funkcí. Crutchfield, Mitchellová a Das (2003) takto s pomocí optimalizační techniky genetických algoritmů vytvořily ECA, který je schopen realizovat jednoduchou klasifikační úlohu spočívající v rozhodnutí, zda je v počáteční konfiguraci zastoupeno více jedniček nebo nul: Když převažují jedničky, nebo nuly, ECA v daném počtu kroků dospěje ve svém stavovém prostoru k pevnému bodu, kdy všechny jeho buňky jsou ve stavu 1 (černá barva), resp. 0 (bílá barva) (viz obrázek 73). Stejným způsobem,



Obrázek 73: *Prostoro-časový diagram zobrazující chování elementárního celulárního automatu (ECA), který dokáže u libovolné počáteční konfigurace v čase t_0 určit, zda v ní převažují jedničky (černá barva), nebo nuly (bílá barva). Odpověď je zakódována v podobě pevného bodu, kdy jsou všechny buňky ECA ve stavu 1 (když ve výchozí konfiguraci převažují jedničky), resp. ve stavu 0 (to když převažují nuly). Pravidlo, kterým se řídí změna stavů jednotlivých buněk, byl vytvořen s pomocí optimalizační techniky genetických algoritmů. (s úpravami převzato z Crutchfield, Mitchell, Das, 2003)*

tj. s pomocí genetických algoritmů, se podařilo vytvořit ECA, který zvládá synchronizační úlohu, která spočívá v tom, že ať už je počáteční konfigurace jakákoli, ECA musí vždy po určitém počtu kroků skončit ve stavu, kdy se stav všech buněk synchronizovaně a pravidelně překlápí mezi nulou a jedničkou (Das, Crutchfield, Mitchell, Hanson, 1995). Vyjádřeno v terminologii teorie dynamických systémů to znamená, že dlouhodobé chování tohoto ECA,

resp. výpočet, který provádí, charakterizuje limitní cyklus, kdy se ECA po stejném počtu kroků vrací do stejného stavu (viz obrázek 74). Jakkoli je velice obtížné uchopit to, jak ECA svůj výpočet konkrétně provádí, určitý vhled do tohoto procesu poskytuje grafické znázornění jeho stavového prostoru, který člověku umožňuje zabývat se ECA na vyšší úrovni celkových vzorců chování a do určité míry ignorovat nepřehlednou změť jednotlivých buněk a dílčích přechodových pravidel. Na základě obdobného principu funguje také řada dalších technik a nástrojů, které jsou předmětem následujícího oddílu.



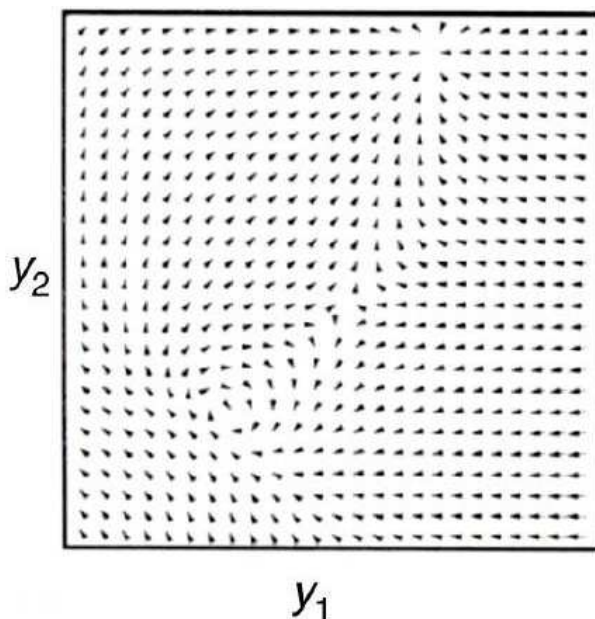
Obrázek 74: Prostoro-časový diagram chování elementárního celulárního automatu (ECA) vyvinutého s pomocí genetických algoritmů za účelem realizaci synchronizační úlohy. Bez ohledu na konkrétní detaily počáteční konfigurace ECA končí v pevně daném počtu kroků ve stavu, kdy se stav všech buněk synchronizovaně a pravidelně překlápí mezi nulou a jedničkou. (s úpravami převzato z Das, Crutchfield, Mitchell, Hanson, 1995)

2.2.1.2 Základní nástroje vizualizace a analýzy chování komplexních dynamických systémů

Vzhledem k vysoce abstraktní povaze matematických rovnic popisujících dynamické systémy se k usnadnění analýzy jejich chování v čase používají různé vizualizační techniky, které jsou založeny na některých konceptech představených v předchozím oddíle. Beer (2000) demonstruje použití některých těchto nástrojů na jednoduchém dvojdimenzionálním dynamickém systému, jehož chování je specifikováno dvěma diferenciálními rovnicemi:

$$\dot{y}_1 = f_1(y_1, y_2) \quad \dot{y}_2 = f_2(y_1, y_2)$$

Tyto dvě diferenciální rovnice lze vizualizovat prostřednictvím dvojrozměrného **vektorového pole** (*vector field*; viz obrázek 75), ve kterém je každému bodu stavového prostoru dynamického systému přisouzen určitý specifický směr a velikost změny systémových proměnných y_1 a y_2 .¹

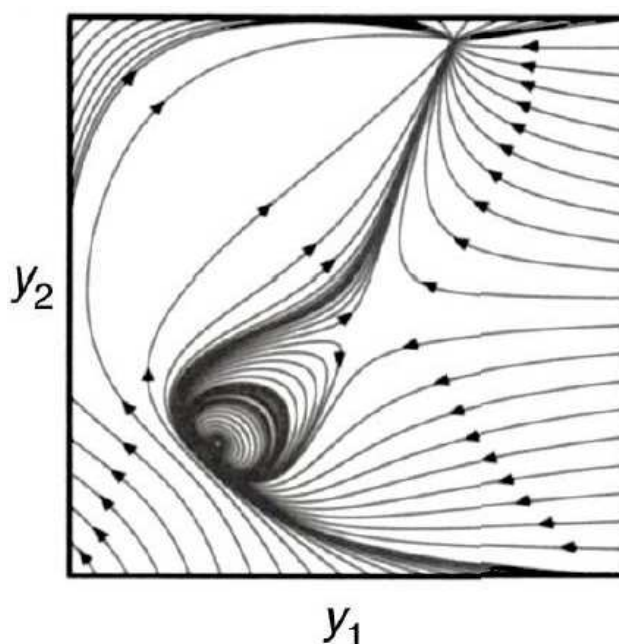


Obrázek 75: Vektorové pole definované dvěma diferenciálními rovnicemi $\dot{y} = f_1(y_1, y_2)$ a $\dot{y} = f_2(y_1, y_2)$. popisujícími chování jednoduchého dvojdimenzionálního dynamického systému v čase. Ve vektorovém poli je každému bodu stavového prostoru dynamického systému přisouzena určitá konkrétní velikost a směr změn dvou stavových proměnných. (převzato z Beer, 2000)

Dalším nástrojem vizualizace chování dynamického systému je **diagram toků** (*flow diagram*), ve kterém jsou zachyceny všechny hlavní trajektorie systému ve stavovém prostoru, které vznikají aplikací vývojového operátoru na různé počáteční stavy systému (viz obrázek 76).

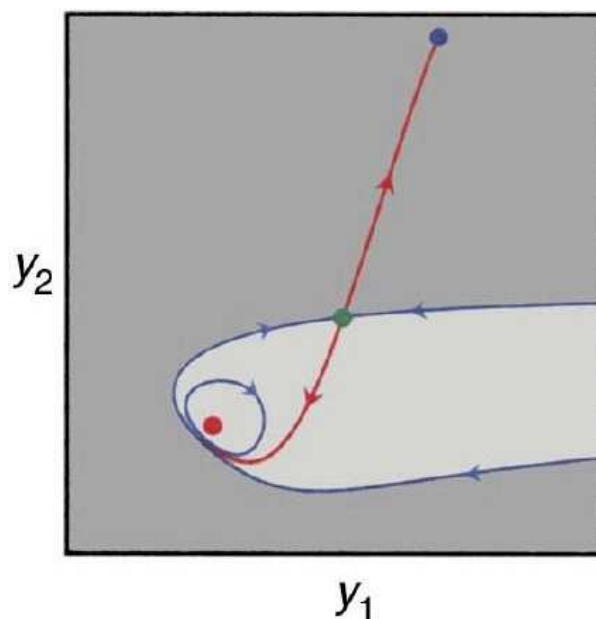
¹ Míru změny hodnoty nějaké proměnné vyjadřuje její derivace. Derivace je obecně poměr, v jakém změna nějaké proměnné y odpovídá změně jiné proměnné x , na které je tato proměnná funkčně závislá. Tento poměr lze symbolicky zapsat jako $\frac{dy}{dx}$; jiný způsob notace používá tečku nad proměnnou $\dot{x} = \frac{dx}{dt}$; tato notace se většinou

používá ve fyzice pro derivování podle proměnné vyjadřující čas (t). Derivace funkce $f(x)$ vyjadřuje rychlost změny této funkce vzhledem k jejímu parametru či parametrům. V případě dvourozměrného grafu funkce $f(x)$ je derivace této funkce v libovolném bodě rovna směrnici tečny tohoto grafu. Derivace je tak v podstatě náklon grafu dané funkce.



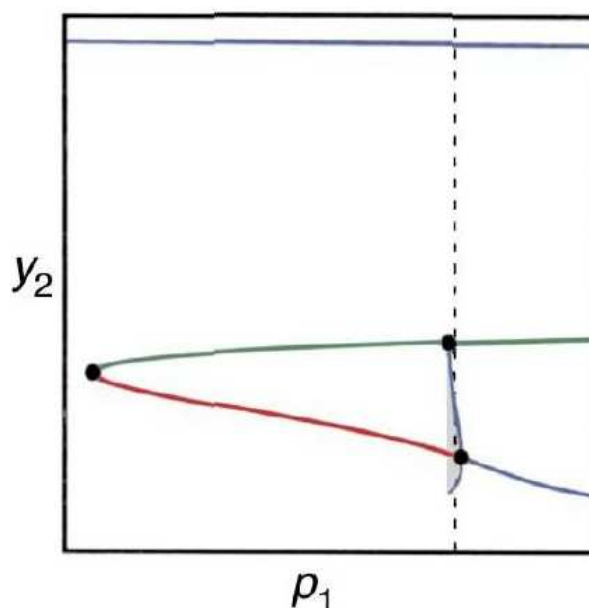
Obrázek 76: Diagram toků zobrazující všechny hlavní vývojové trajektorie jednoduchého dvojdimenzionálního dynamického systému v jeho stavovém prostoru. (převzato z Beer, 2000)

K přehlednému grafickému zachycení dlouhodobého chování dynamického systému slouží **fázový digram** (*phase portrait*). Jak již bylo uvedeno v předchozím oddíle, většina dynamických systémů má tendenci se v průběhu času vyskytovat jen v několika málo oblastech svého stavového prostoru, tedy jen v omezeném souboru všech možných stavů. V této souvislosti se hovoří o tzv. *limitních množinách* (nebo také *atraktorech*), které mohou mít podobu *limitních* (nebo také *rovnovážných*) *bodů* (kdy stav dynamického systému zůstává v čase stabilní a nijak se nemění), nebo *limitních cyklů*, kdy se dynamický systém v určitém počtu kroků pravidelně vrací do stejného stavu. Tyto limitní množiny přitom mohou být stabilní, nebo nestabilní. V případě stabilního atraktoru daná limitní množina přitahuje všechny okolo jdoucí trajektorie; v případě nestabilního atraktoru naopak sebemenší odchylka od dané limitní množiny vede k odchýlení trajektorie do jiných oblastí stavového prostoru (*repelory*). Vedle toho existují také limitní množiny, které jsou obecně nestabilní, avšak v některých specifických směrech mohou být stabilní (tzv. *sedlovité limitní množiny*, *saddle limit sets*). V souvislosti s těmito vlastnostmi limitních množin se hovoří také o *basins of attraction*, tedy o množině bodů nacházejících se v okolí atraktorů, které jako gravitační pole stahují okolo jdoucí trajektorie k dané limitní množině stavů. Všechny tyto objekty – limitní body, limitní cykly, repelory, basins of attraction -, které charakterizují dlouhodobé chování dynamického systému, zachycuje fázový diagram (viz obrázek 77).



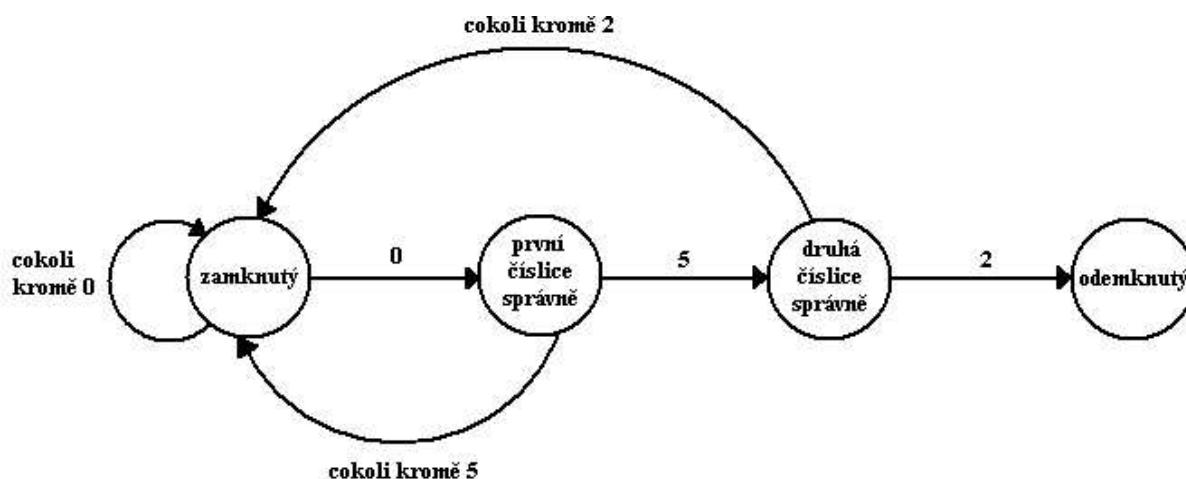
Obrázek 77: Fázový diagram zachycující limitní množiny a basins of attraction jednoduchého dvojdimenzionálního dynamického systému. Limitní body jsou v diagramu znázorněny jako tečky a limitní cyklus jako oválná trajektorie; stabilní limitní množiny mají modrou barvu, nestabilní barvu červenou a sedlové zelenou. Světle šedou barvou je vyznačena basin of attraction limitního cyklu; tmavě šedou pak basin of attraction stabilního limitního bodu. Modrá a červená dráha představují stabilní, resp. nestabilní směry oblasti stavového prostoru vedoucí k sedlovému limitnímu bodu, resp. od něj. (převzato z Beer, 2000)

Důležitým nástrojem analýzy chování dynamického systému je **diagram bifurkace** (*bifurcation diagram*), který zachycuje změnu „topologie“ fázového diagramu v závislosti na změnách hodnot řídicích parametrů dynamického systému. Topologie fázového diagramu se většinou mění hladce a pozvolna, občas však v tzv. bifurkačních bodech (na obrázku 78 označených černými body) dochází k prudké a radikální změně tvaru „kopců“ a „údolí“ fázového diagramu a to navzdory jen nepatrným změnám v hodnotách řídicích parametrů. Na obrázku 78 takto můžeme vidět diagram bifurkace pro náš jednoduchý dvojdimenzionální dynamický systém, na kterém je znázorněna změna jeho dlouhodobého chování v závislosti na změně řídicího parametru p_1 : Při nízkých hodnotách p_1 je chování systému charakterizováno pouze jedním stabilním limitním bodem (horní modrá přímka). Jak hodnota p_1 narůstá, objevuje se další dvojice limitních bodů, z nichž jeden je nestabilní (červená křivka) a druhý sedlový (zelená křivka). S dalším nárůstem p_1 se objevuje také stabilní limitní cyklus (šedá oblast ohraničená modrou křivkou); tato situace odpovídá stavu zachycenému ve fázovém diagramu na obrázku 78. Při dalším nárůstu hodnoty parametru p_1 limitní cyklus mizí a zůstávají pouze dva stabilní limitní body, které od sebe odděluje jeden sedlový limitní bod.



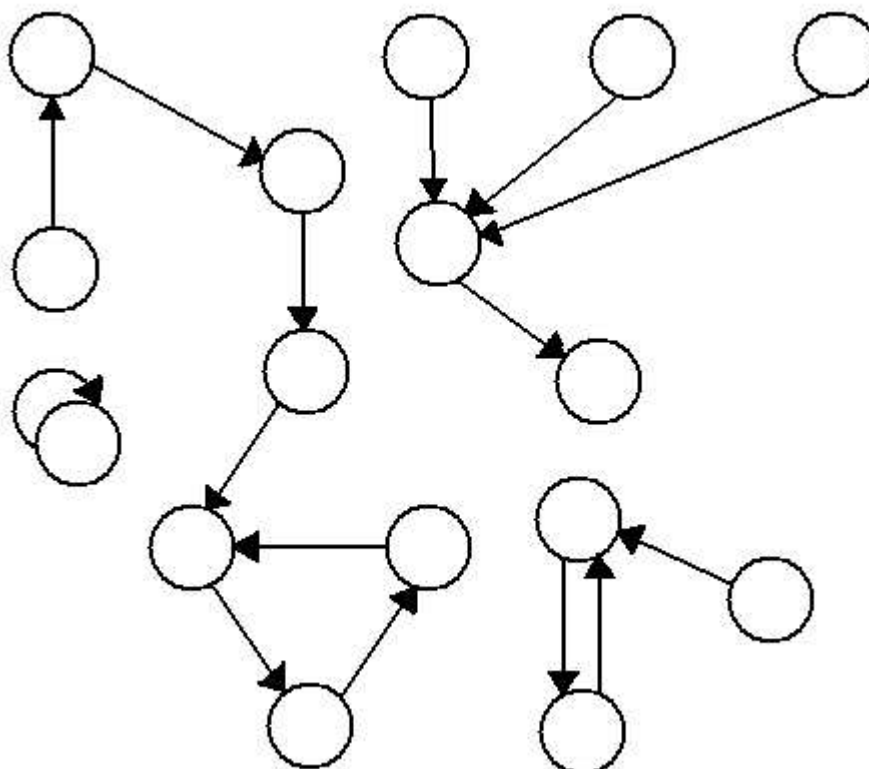
Obrázek 78: Bifurkační diagram zachycující změny topologie fázového diagramu v závislosti na změnách řídicího parametru p_1 ; více viz hlavní text. (převzato z Beer, 2000)

K analýze chování dynamických systémů se také často používá tzv. **diagram stavových přechodů** (*state transition diagram*), který je vhodný zejména pro analýzu diskretních dynamických systémů. Tento analytický nástroj je založen na pojmání dynamického systému jako druhu konečného automatu, který se v každém okamžiku nachází v jednom z konečné množiny stavů a jehož stav se mění v závislosti na jeho aktuálním stavu a přijímaném vstupu (viz také s. 53-54). Diagram stavových přechodů (dále DSP) pak nabízí přehledný způsob vizualizace chování takového konečného automatu v podobě grafického znázornění struktury jeho stavového prostoru, resp. přechodů mezi jeho jednotlivými stavy. DSP se skládá ze dvou základních prvků - z kruhů, které reprezentují jednotlivé stavy systému, a z šipek, které reprezentují přechody mezi jednotlivými stavy systému. Na obrázku 79 jsou takto znázorněny



Obrázek 79: Diagram stavových přechodů zámku s kombinací 0-5-2. (převzato z Hillis, 2003, s. 46)

přechody mezi čtyřmi možnými stavy jednoduchého konečného automatu - kombinačního zámku, který se otevře pouze při zadání posloupnosti čísel 0-5-2. Na obrázku 80 je pak DSP znázorňující strukturu stavového prostoru dynamického systému, který se může nacházet v šestnácti různých stavech a jehož chování charakterizují – jak je dobře patrné z DSP – čtyři limitní množiny (dva pevné body a dva limitní cykly); další příklad DSP je možné nalézt na straně 184 na obrázku 102, kde je DSP jednoduché binární fuzzy kognitivní mapy.



Obrázek 80: Diagram stavových přechodů dynamického systému, který se může nacházet v šestnácti různých stavech. V diagramu jsou na první pohled patrné čtyři limitní množiny – dva pevné body a dva limitní cykly.

2.2.1.3 Některé další nástroje zkoumání komplexních dynamických systémů

Základní funkcí lidské mysli jakožto řídicího systému je využívání informací o vnějším prostředí k vyhýbání se hloupým tahům a k vytváření chytrých kroků, které redukuje pravděpodobnost, že se člověk v rozlehlém a složitém světě vydá špatným směrem. Královskou cestou, jak snížit nejistotu ohledně svého dalšího kroku, je schopnost vidět dále dopředu v čase, tj. schopnost predikce: Při rozhodování se mezi různými alternativními postupy či řešeními člověk většinou jedná na základě svých mentálních modelů, které jsou vnitřní zjednodušenou reprezentací vnějšího světa. Díky tomu, že tyto mentální modely

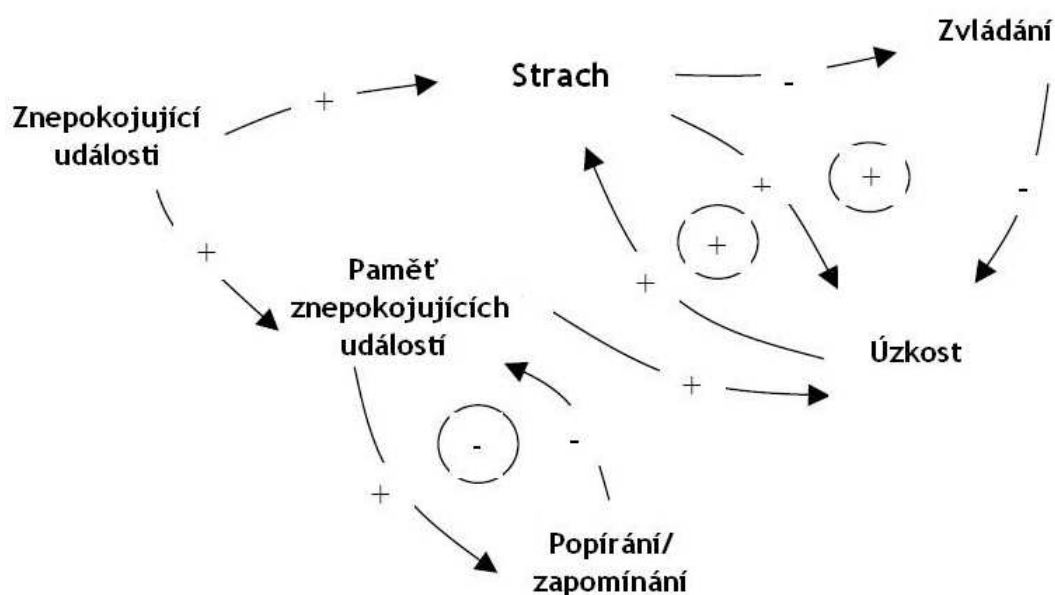
obsahují mnoho informací o vnějším světě a o jeho zákonitostech, má člověk možnost v tomto vnitřním prostředí bezpečně provádět pokusy a simulace spojené s mentálním testováním možných budoucích důsledků svých rozhodnutí a „*nechat [tak]*“, řečeno slovy filozofa *Karla Raimunda Poppera*, „*namísto [sebe] umírat [své] hypotézy*“ (citováno podle Dennett, 2004b, s. 86). Tato schopnost mentální simulace možných důsledků alternativních tahů a kroků představuje zásadní vylepšení základního skinnerovského designu mysli, jejíž řídicí činnost je založena čistě na principu podmiňování: „*Skinnerovské podmiňování [v rámci namáhavého procesu pokusu–a–omylu] je dobrá věc, pokud nejste zabiti jednou z vašich raných chyb. Lepší systém zahrnuje předvýběr mezi všemi možnými způsoby chování nebo činy, takže ty skutečně hloupé tahy jsou vyplety dříve, než jsou riskovány ve „skutečném životě“. [...] Na rozdíl od pouze skinnerovských tvorů, z nichž mnozí přežijí jenom proto, že udělají šťastné první tahy, my lidé, ale nejenom my, přežíváme proto, že jsme dost chytrí na to, abychom udělali ty první tahy lepší než jen náhodné. Máme, samozřejmě, štěstí, že jsme chytrí, to je však lepší než jenom mít štěstí.*“ (Dennett, 2004b, s. 86)

Naneštěstí jsou možnosti mentálních modelů co se týče jejich schopnosti věrně reprezentovat vnější prostředí dosti omezené. To platí zvláště v souvislosti s mentálním modelováním komplexních dynamických systémů, pro které je charakteristická komplexní cirkulární kauzalita a nelineární vztahy mezi jednotlivými prvky systému. Důsledkem toho je pak to, že člověk vnější svět ve své mysli často reprezentuje nepřesně, takže nepřesná jsou potom i jeho očekávání ohledně toho, jak se bude vnější svět chovat nebo jaké postupy a strategie povedou ke kýženému cíli, což může mít v řadě různých oblastí lidské činnosti mnoho nedozírných následků. Proto vznikla řada různých metod a postupů vycházejících z poznatků teorie dynamických systémů, které člověku analýzu chování komplexních dynamických systémů usnadňují.

2.2.1.3.1 Příčinné smyčkové diagramy

Jednou takovou jednoduchou metodou jsou **příčinné smyčkové diagramy** (dále PSD), které umožňují přehledným způsobem reprezentovat prvky dynamického systému a komplexní kauzální vztahy mezi nimi. PSD sestávají z dynamických proměnných, které jsou mezi sebou propojeny prostřednictvím šipek, které specifikují směr kauzálního působení mezi jednotlivými prvky systému. Spojení mezi proměnnými přitom může být buď pozitivní, nebo negativní, to podle toho, zda působící proměnná posiluje, nebo naopak oslabuje vybranou stavovou vlastnost ovlivňované proměnné. Jako příklad PSD zde můžeme uvést digram

intrapsychické dynamiky strachu u jednotlivce nebo skupiny jednotlivců (Pauková, 2005): V příčinném diagramu (viz obrázek 81) jsou zachyceny následující prvky: pocit strachu, znepokojující události, paměť znepokojujících událostí, popírání a zapomínání vzpomínek na znepokojující události, úzkost a zvládání pocitů strachu a úzkosti (tj. vyrovnávání se s těmito emocemi). Zobrazený systém obsahuje tři uzavřené zpětnovazební smyčky, dvě pozitivní



Obrázek 81: Příčinný smyčkový diagram intrapsychické dynamiky strachu u jednotlivce nebo skupiny jednotlivců v reakci na přijetí zpráv o znepokojujících událostech. (s mírnými úpravami převzato z Pauková, 2005)

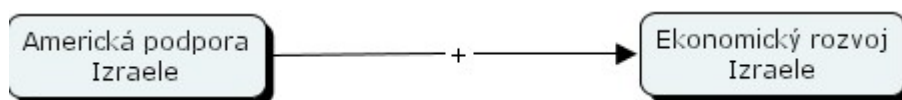
pozitivní a jednu negativní. První pozitivní zpětná vazba existuje mezi strachem a úzkostí, tzn. že tyto dvě proměnné se navzájem potencují, což by v případě absence nebo oslabení stabilizační negativní zpětné vazby v systému strachu mohlo vést k exponenciálnímu růstu strachu a úzkosti. Charakter pozitivní zpětné vazby má také vztah mezi strachem, zvládáním a úzkostí: S narůstajícím strachem se snižuje schopnost se vyrovnávat s negativními emocemi strachu a úzkosti, což vede k nárůstu úzkosti a nárůst úzkosti pak zase vede k nárůstu strachu. Stabilizační roli v zobrazeném systému strachu hraje negativní zpětná vazba mezi pamětí znepokojujících událostí a mechanismem zapomínání a potlačování nepříjemných vzpomínek: S narůstajícím počtem vzpomínek na znepokojující události se automaticky zvyšuje počet zapomenutých nebo potlačených vzpomínek, což v konečném důsledku vede ke snížení počtu vědomých vzpomínek na znepokojující události. Z příčinného smyčkového diagramu je na první pohled zřejmé, že zobrazený systém strachu je díky své specifické struktuře kauzálních

vztahů extrémně nestabilní a náchylný ke katastrofálním selháním (v podobě nekontrolovatelného nárůstu strachu a úzkosti), která jsou typická pro všechny systémy, u nichž došlo ke zhroucení negativních (stabilizujících, kompenzačních) zpětných vazeb.

2.2.1.3.2 Kognitivní mapy¹

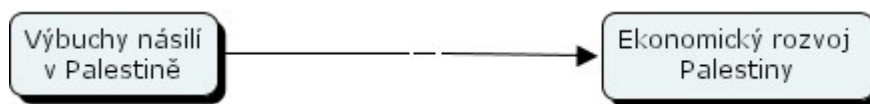
Obdobnou funkci a strukturu jako PSD mají také tzv. **kognitivní mapy**. Pojem *kognitivní mapa* do psychologie zavedl *Edward C. Tolman*, jeden z čelních představitelů neobehaviorismu, který kognitivní mapou rozuměl vnitřní mentální mapu okolního prostředí, kterou organismus využívá k orientaci a k plánování činnosti v tomto prostředí. K mapování a modelování příčinných a implikačních vztahů mezi různými jevy a událostmi se však používají kognitivní mapy ve významu, jaký jim dal americký politolog *R. Axelrod* (1976), který kognitivní mapy použil k modelování myšlenkových schémat politických vztahů v mysli člověka.

Takto chápaná kognitivní mapa představuje systém, který se skládá ze souboru pojmů a ze souboru příčinných vztahů mezi těmito pojmy, přičemž každý z pojmů může prostřednictvím kauzálních vztahů pozitivně (*excitačně*) nebo negativně (*inhibičně*) ovlivňovat ostatní pojmy, se kterými je takto příčinně propojen. Kognitivní mapa má podobu ohodnoceného orientovaného grafu, který sestává z uzlových bodů, které odpovídají jednotlivým pojmům, a z šipek odpovídajících kauzálním vztahům mezi pojmy; tyto šipky směřují vždy od „příčinného pojmu“ směrem k „důsledkovému pojmu“ a navíc jsou vždy ohodnoceny buď znaménkem +, nebo –, to podle toho, zda je daný příčinný vztah pozitivní (*excitační*), nebo negativní (*inhibiční*). Pro ilustraci je na obrázku 82 zachycen pozitivní kauzální vztah mezi americkou podporou Izraele a jeho ekonomickým rozvojem a na obrázku 83 pak negativní příčinný vztah mezi výbuchy násilí v Palestině a její ekonomickou prosperitou.



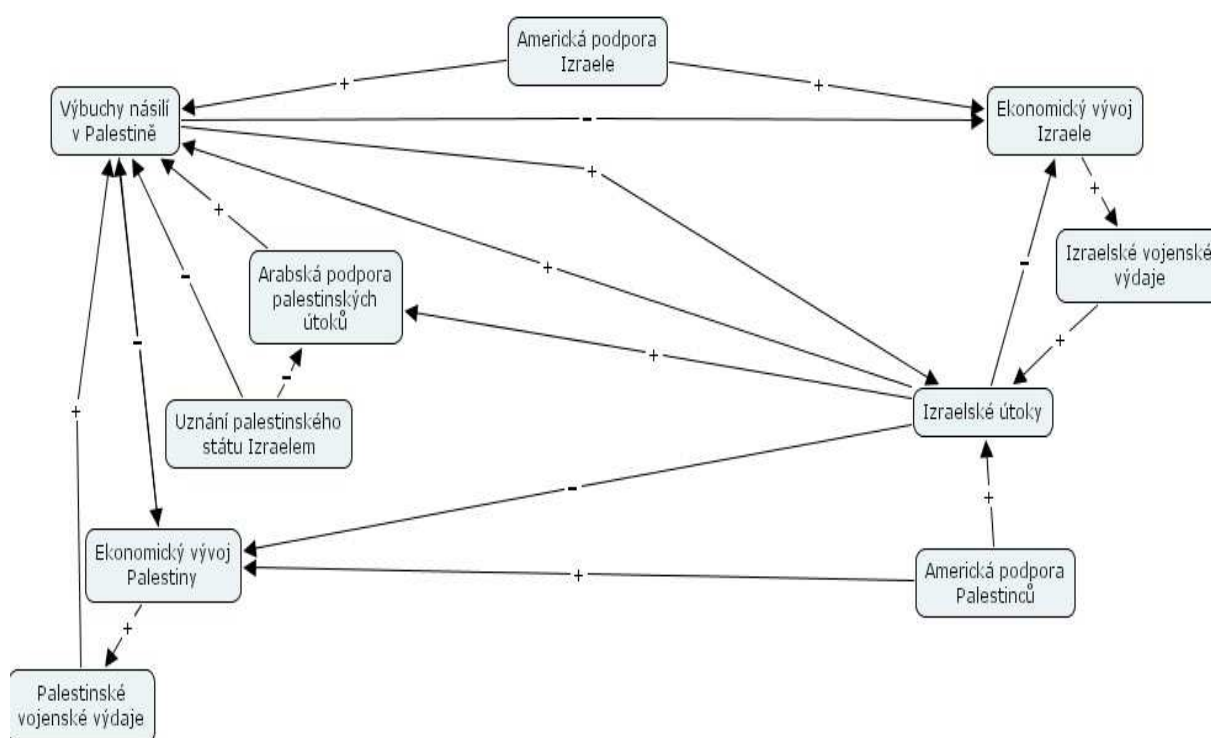
Obrázek 82: Jednoduchá kognitivní mapa zachycující pozitivní příčinnou vazbu mezi americkou podporou Izraele a jeho hospodářskou prosperitou: Čím větší bude podpora Izraele ze strany USA, tím větší bude i ekonomická prosperita Izraele.

¹ Tento oddíl je mírně přepracovanou verzí stejnojmenné kapitoly z mé práce *Nástroje mapování a modelování poznatkových schémat* (Stehlík, 2007).



Obrázek 83: Jednoduchá kognitivní mapa zachycující negativní příčinnou vazbu mezi násilnostmi v Palestině a ekonomickým vývojem v Palestině: Čím větší nepokoje budou v Palestině, tím horší bude v Palestině hospodářský vývoj.

Tímto způsobem pak lze sestavit kognitivní mapu, která bude zachycovat jeden z mnoha možných pohledů na konflikt mezi Izraelem a Palestinci (viz obrázek 84), který představuje velice složitou situaci zahrnující velké množství různých faktorů a kauzálních vztahů mezi těmito faktory s řadou zpětnovazebních smyček. Máme-li vytvořenu takovouto kognitivní

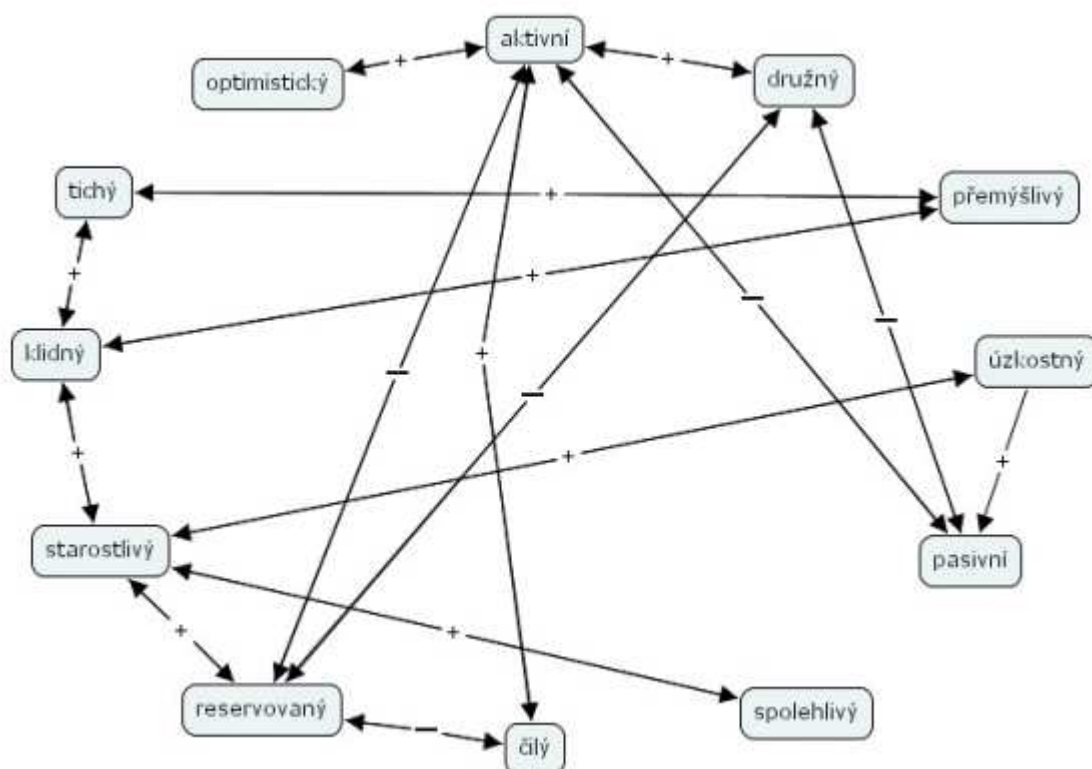


Obrázek 84: Kognitivní mapa politické situace na středním Východě. (s mírnými úpravami převzato z <http://www.ochoadeaspuru.com/fuzcogmap/middleeast.php>)

mapu, můžeme začít s modelováním různých možných scénářů vývoje politické situace na středním Východě; jedná se de facto o hru na to, co by se stalo, kdyby...: *Jestliže USA budou podporovat Izrael, Izraeli se bude po ekonomické stránce dobře dařit, díky čemuž bude schopen navýšit svůj vojenský rozpočet. To bude mít za důsledek větší tendenci Izraele řešit konflikty s Palestinci prostřednictvím vojenských útoků, které jen vyprovokují podobnou reakci na straně Palestinců, kteří v tomto najdou silnou podporu okolních arabských států.*

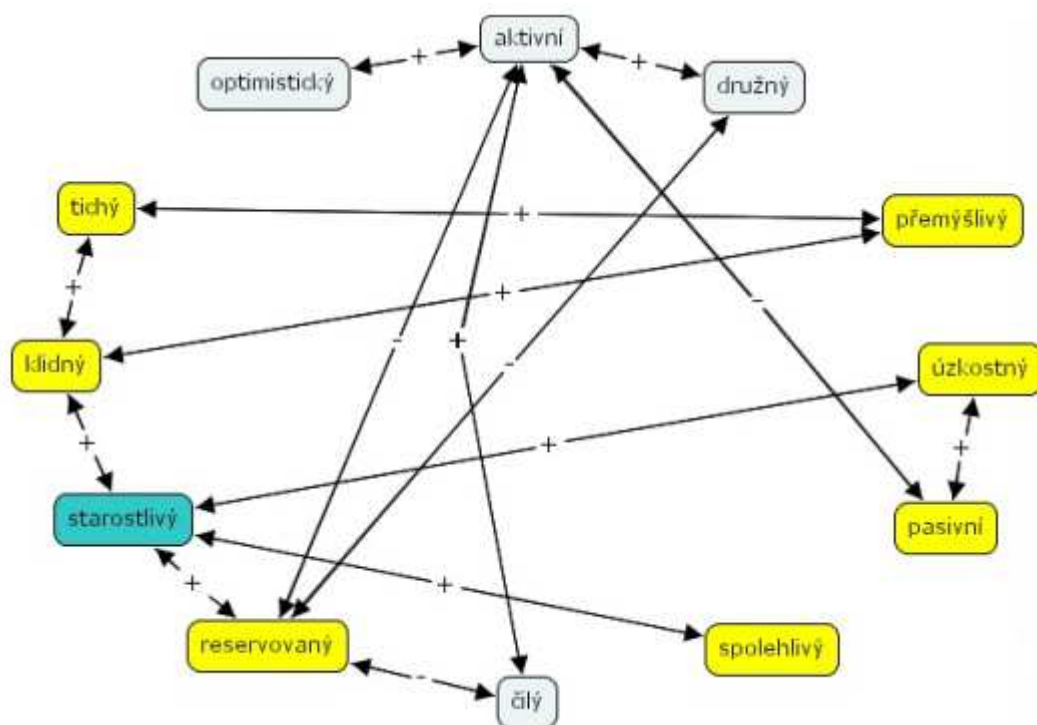
Tato atmosféra povede k ekonomické stagnaci či propadu, který znemožní Izraeli štedře financovat armádu a její vojenské akce na palestinském území, což může mít za následek větší ochotu Izraele řešit konflikt s Palestinci mírovou cestou...

Kognitivní mapy lze používat nejen k modelování politických vztahů, ale obecně k modelování jakýchkoli komplexních systémů, v nichž mezi jednotlivými prvky existuje spletité předivo různých kauzálních a implikačních vztahů a souvislostí, které lze vyjádřit v podobě pravidel typu JESTLIŽE-PAK. Takto lze například modelovat různá **sociální schémata**, která představují určitým způsobem organizované systémy pojmů, poznatků a přesvědčení o lidech, předmětech, událostech a situacích, které člověku umožňují identifikovat významy jednotlivých událostí a jevů, zorientovat se v sociálních situacích a anticipovat chování předmětů a druhých lidí i důsledky svých vlastních akcí (Nakonečný, 1999, s. 96). Na obrázku 85 je zachycena kognitivní mapa **implicitní (laické) teorie osobnosti**, kterou si lidé vytvářejí o tom, jak spolu souvisejí jednotlivé rysy osobnosti extroverta, resp. introverta: Takovýmto způsobem propojené osobnostní rysy člověku poskytují kontext k interpretaci vstupních dat a umožňují usuzovat i na charakteristiky, které



Obrázek 85: Kognitivní mapa zachycující implicitní teorii osobnosti (introverze/extroverze).

nejsou ve vstupních datech bezprostředně přítomné.¹ Takže když člověk s takovouto implicitní teorií osobnosti obdrží explicitní informaci o tom, že osoba, s níž se má dnes večer setkat, je *starostlivá*, potom tato vstupní informace vede k aktivaci či inhibici uzlů celé sítě, které představují jednotlivé charakteristiky zapojené v dané implicitní teorii osobnosti; v tomto případě se aktivují uzly *klidnosti*, *úzkostnosti*, *spolehlivosti*, *přemýšlivosti*, *tichosti*, *rezervovanosti*, *pasivnosti* a inhibují se uzly *čilosti*, *družnosti*, *aktivnosti* a *optimističnosti* (viz obrázek 86). S takto nastavenou (zapojenou) implicitní teorií osobnosti a na základě dané vstupní informace tedy bude člověk předpokládat, že dnes večer se nejspíš setká s člověkem, pro něhož jsou charakteristické vlastnosti introverta, a že je tedy potřeba se na jednání s takovým člověkem odpovídajícím způsobem připravit.



Obrázek 86: Žlutě označené uzlové body představují pojmy aktivované v reakci na poskytnutou informaci o tom, že daný člověk je starostlivý (modrý pojmový uzel).

¹ Zpracovávání informací prostřednictvím různých schémat (tzv. schematické zpracovávání informací) probíhá tak, že člověk ve své paměti vyhledává v minulosti vytvořená schémata, která jsou se vstupními informacemi nejkonzistentnější, a na základě těchto schémat potom vstupní informace interpretuje. Tento způsob zpracovávání informací je zdrojem kognitivní ekonomičnosti, neboť člověku umožňuje rychle, automaticky a efektivně uspořádat a zpracovávat velké množství informací, a to i v situacích, které jsou zatíženy informačním šumem. Člověk takto například není nucen si všimnout nebo si pamatovat každý detail nějakého nového jevu, se kterým se potká, ale stačí mu, když si všimne, že tento nový jev se v několika málo nejdůležitějších charakteristikách podobá jednomu z několika v minulosti nabytých schémat (např. introvert), představujících uspořádané soubory poznatků a přesvědčení o jednotlivých instancích dané třídy jevů (např. uzavřenost, nekomunikativnost, nespolečenskost, úzkostnost, rezervovanost, přemýšlivost, tichost...), která jsou všechna okamžitě člověku automaticky přístupná, jakmile je rozpoznána základní podobnost jevu s daným schématem.

2.2.1.3.3 Dynamické modelování

Jakkoli jsou příčinné smyčkové diagramy a kognitivní mapy užitečné k přehlednému zachycení hypotéz ohledně příčin dynamiky systému nebo k zobrazení myšlenkových schémat a mentálních modelů, které si lidé vytvářejí o struktuře kauzálních a implikačních vztahů mezi prvky nějakého systému, možnosti využití těchto nástrojů ke skutečnému modelování dynamických systémů jsou dosti omezené, neboť pracují jen s velice vágně definovanými proměnnými a vztahy mezi nimi. Tzn. že při zachycení kauzální struktury nějakého dynamického systému prostřednictvím PSD nebo kognitivní mapy člověk získá jen velice obecnou představu o dynamice chování zkoumaného systému v čase, která je důsledkem jeho kauzální struktury.

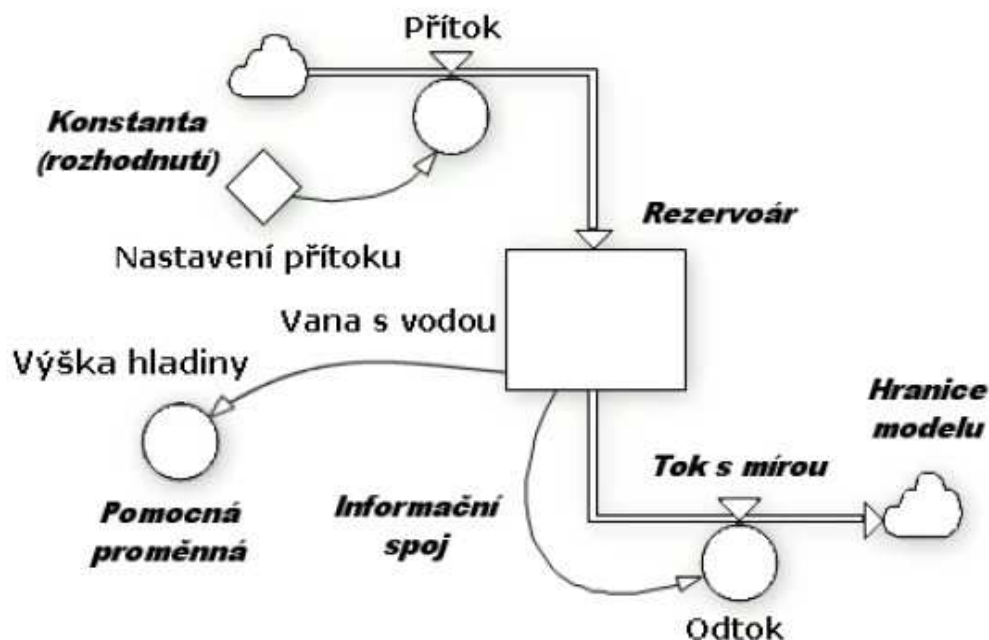
K precizaci zkoumání dynamických systémů je nutné jednotlivé systémové prvky a vztahy mezi nimi přesně definovat a to nejlépe v jazyce matematiky. Takové matematické modelování dynamických systémů má obvykle podobu hledání a formulování souboru poměrně složitých diferenciálních rovnic, které by dokázaly zachytit dynamiku chování sledovaného systému. Naštěstí na trhu existuje několik softwarových produktů (například *Powersim Studio*, *iThink/Stella*, *Vensim*), které výzkumníkovi poskytují uživatelsky vstřícné simulační prostředí, ve kterém při modelování dynamických systémů vystačí de facto pouze se znalostmi základních aritmetických operací.

Základními prvky těchto modelů dynamických systémů jsou **akumulace** (také *hladiny* nebo *stavy*) a **toky** (Šusta, Neumaierová, 2004, s. 18). Akumulace je určitým druhem rezervoáru, ve kterém dochází ke hromadění nějaké proměnné. Hromadit se přitom může cokoli, co lze kvantifikovat: zboží, peníze, zásoby, voda, prach, emoce, stres atd. Každá akumulace má v daném okamžiku určitou konkrétní hodnotu. Hodnota akumulace se v čase mění a to prostřednictvím činnosti toků, které zajišťují přítok a odtok z akumulací podobně jako vodovodní kohoutek a odpadní roura zajišťují přítok a odtok vody z vany nebo umyvadla. Důležitou vlastností akumulací je to, že mají svojí vlastní paměť, což se projevuje



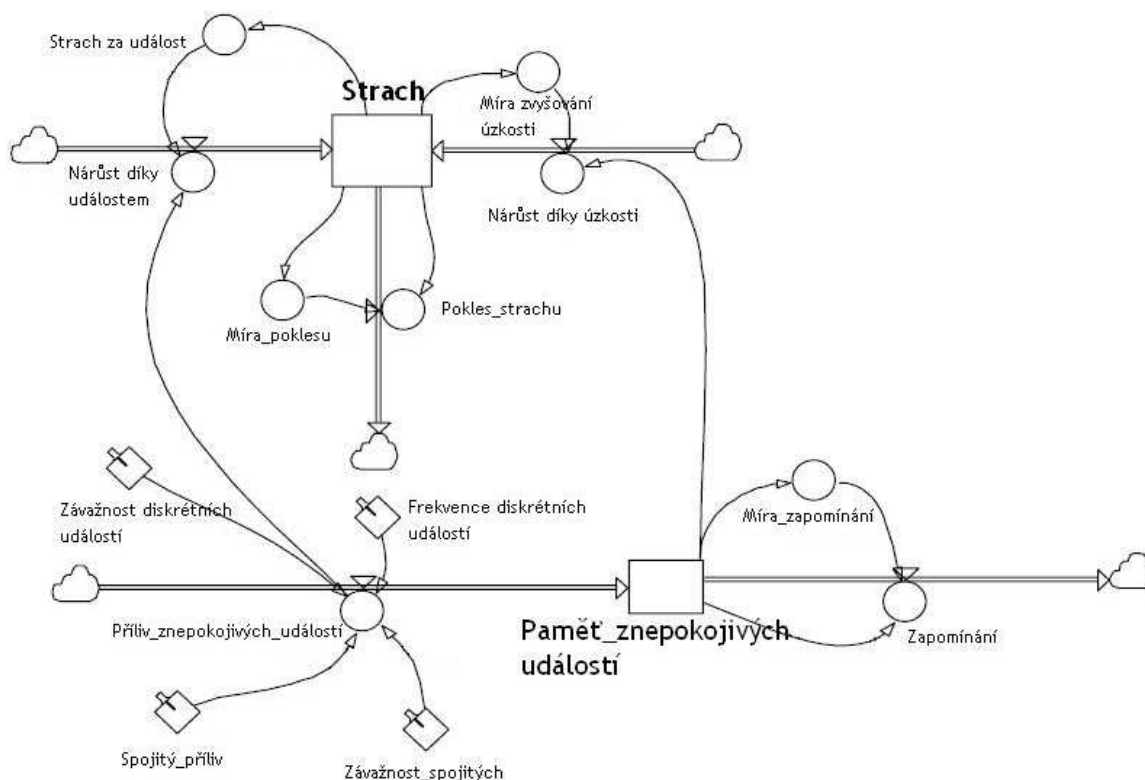
Obrázek 87: Variace jednoduché akumulací a tokové struktury. (převzato z Šusta, Neumaierová, 2004, s. 20; použitá symbolika pochází ze simulačního prostředí Powersim Studio)

tím, že se nemění okamžitě, ale vždy s určitým zpožděním. Vedle akumulací a toků se v rámci modelování dynamických systémů používají další objekty; jedná se především o **konstanty** a **pomocné proměnné**. Konstanty – jak už napovídá jejich samotný název – jsou proměnné, které se v průběhu simulace nijak nemění a mají stálou hodnotu. Konstanty se obvykle používají k znázornění exogenních proměnných (tj. prvků, které se nacházejí mimo hranice systému) nebo takových proměnných, o kterých se předpokládá, že se v průběhu času nijak významně nemění. Pomocné proměnné se pak používají ke specifikaci některých informací o akumulacích a tocích. Na obrázku 88 je na jednoduchém systému – modelu vany s přítokem a odtokem – demonstrováno použití některých základních prvků diagramu stavů a toků (rezervoár, tok s mírou, pomocná, proměnná, konstanta, informační spoj) v symbolice simulačního prostředí Powersim Studio. Na obrázku 89 je pak diagram stavů a toků, který vychází z příčinného smyčkového diagramu zachycujícího intrapsychickou dynamiku strachu u jednotlivce nebo skupiny jednotlivců (viz obrázek 81); v příloze na konci práce lze nalézt také rovnice definující jednotlivé prvky diagramu; případný zájemce tak má možnost - po stažení 30-denní trial verze programu Powersim Studio¹ - experimentovat s dynamikou tohoto systému, případně celý systém modifikovat a doplnit ho o nové proměnné a vztahy, které by model v jeho chování učinily realističtější.



Obrázek 88: Diagram stavů a toků jednoduchého systému – modelu vany s přítokem a odtokem. (převzato z www.proverbs.cz)

¹ Bližší informace o programu Powersim Studio lze nalézt na adrese www.proverbs.cz nebo přímo na adrese výrobce www.powersim.com. Právě na druhé jmenované adrese je možné si stáhnout 30-denní, plně funkční trial verzi programu Powersim Studio.



Obrázek 89: Diagram stavů a toků modelu dynamiky strachu jednotlivce či skupiny jednotlivců po přílivu znepokojivých zpráv. Znepokojivé události vyvolávají u člověka nebo skupiny lidí emoční reakci a mají tak přímý vliv na nárůst strachu. Současně se tyto události ukládají do paměti znepokojivých událostí, která hraje roli mediátoru nárůstu strachu z důvodu existence dřívějších úzkostných stavů. Příčinná struktura obsahuje dva odtoky, které umožňují postupné odbourávání strachu a snižování hladiny paměti znepokojivých událostí. V příloze A jsou uvedeny rovnice definující jednotlivé prvky tohoto diagramu. (převzato z www.proverbs.cz)

2.2.1.3.4 Fuzzy kognitivní mapy¹

Jinou metodou, jak precizovat modelování komplexních dynamických systémů, jsou **fuzzy kognitivní mapy** (*fuzzy cognitive maps*). Hlavním přínosem běžných kognitivních map je to, že dokáží přehledným a uživatelsky vstřícným způsobem reprezentovat poznatky o různých typech příčinně-důsledkových vztahů a souvislostí mezi prvky nějakého dynamického systému. To, co však běžné kognitivní mapy nezohledňují, je to, že zákonitosti chování komplexních dynamických systémů a stejně tak lidské poznatky o těchto zákonitostech jsou do značné míry zatíženy **neurčitostí** a **nejistotou**, což se v praxi projevuje například tím, že lidé – včetně expertů - ve svém každodenním usuzování o nejrůznějších komplexních dynamických systémech nepoužívají formální modely s přesně nadefinovanými proměnnými

¹ Tento oddíl je mírně přepracovanou verzí stejnojmenné kapitoly z mé práce *Nástroje mapování a modelování poznatkových schémat* (Stehlík, 2007).

a s přesně nadefinovanými vztahy mezi těmito proměnnými, ale spoléhají se na více či méně **vágně formulované poznatky** typu *Přemýšlivý člověk bývá většinou tichý a málomluvný, Dlouhodobě frustrovaný jedinec má často silnou tendenci jednat agresivně* či *Když fouká silný vítr zprava, je většinou třeba při míření a při střelbě trochu předsadit doprava...* Kdo všechno by však spadl do kategorie *přemýšlivý člověk*? Všichni, kdo alespoň jednou denně přemýšlí o „nesmrtelnosti chrousta“, nebo ti, co tak činí minimálně dvakrát denně? Koho by bylo možné označit za *dlouhodobě frustrovaného jedince*? Člověka frustrovaného jeden týden, několik týdnů, jeden měsíc nebo několik měsíců? Kdy lze o *větru* říci, že je *silný*? Při 50, nebo až při 51 kilometrové rychlosti za hodinu? Všechny výše uvedené pojmy jsou neurčité a vágní. Neurčité, resp. nejisté jsou rovněž postulované vztahy mezi těmito pojmy. Co to znamená *většinou* nebo *často*? Šest, nebo osm případů z deseti? Jak silné musí být vnitřní puzení, abychom mohli mluvit o *silné tendenci* k něčemu? Právě s tímto typem neurčitosti obsažené v chování komplexních dynamických systémů a v lidských poznatcích o nich dokáží pracovat fuzzy kognitivní mapy - **nástroj soft-computingu, který je kombinací kognitivních map, fuzzy logiky a umělých neuronových sítí**. Avšak dříve než si ukážeme možnosti využití fuzzy kognitivních map při modelování komplexních dynamických systémů, je potřeba zde zmínit alespoň některé základní údaje o jedné z hlavních komponent fuzzy kognitivních map, totiž o fuzzy logice.

Podle amerického psycholingvisty *Stevena Pinkera* (2003, s. 339-345) lidská mysl vytváří dva základní druhy kategorií, a sice **klasické kategorie** a **fuzzy kategorie**. Pro klasické kategorie je charakteristické, že o tom, který předmět spadá do dané kategorie, rozhoduje to, zda daný předmět vyhovuje či nevyhovuje jednoznačné definici příslušné kategorie (např. *Liché číslo je celé číslo, po kterém zůstane zbytek, jestliže se vydělí dvěma, Babička je matka jednoho ze dvou rodičů* nebo *Starý mládenec je neženatý dospělý muž...*), přičemž o každém objektu lze na základě takovéto definice jednoznačně říci, zda do dané kategorie patří, či nikoli. **U většiny kategorií však není možné jednoznačně určit soubor podmínek členství.** Patří-li do definice ptáka schopnost létat, co potom s takovým tučňákem nebo pštrosem? Židlička vůbec nemusí mít čtyři nohy a nemusí mít ani opěradlo či sedadlo. Je papež starý mládenec, jestliže je starý mládenec definován jako neženatý dospělý muž? Místo přesné definice se u většiny běžných kategorií uplatňuje spíše tzv. **rodinná podobnost a překrývání se rysů**, kdy si jsou členové dané kategorie v něčem podobní a v něčem se naopak liší: Takto je např. zelenina většinou zelená, ale mrkev zelená není; zelenina je v syrovém stavu většinou křupavá, ale pro špenát to neplatí. To má za důsledek to, že **hranice takovýchto kategorií jsou neostré a mlhavé (fuzzy)**, takže si pak lidé například nejsou příliš jistí, zda mají petržel,

česnek nebo mořské řasy považovat za zeleninu, či nikoli. Neostrost fuzzy kategorií se projevuje rovněž v tom, že **některé objekty lidé považují za lepší příklad dané kategorie než objekty jiné**: Tak například většina lidí bude vrabce považovat za daleko lepší příklad ptáka než tučňáka. Tito nejlepší představitelé dané kategorie se nazývají **prototypy** a v mysli lidí shrnují vlastnosti celé kategorie.

Toto rozlišování mezi lepšími a horšími příklady dané kategorie lze dobře zachytit v rámci teorie fuzzy množin, která stojí v základech **multivalentního (vícehodnotového) kalkulu fuzzy logiky** a která narozdíl od **booleovské bivalentní (dvouhodnotové) logiky** připouští **částečnou příslušnost do nějaké množiny**, takže fuzzy množina je pak tvořena prvky, z nichž každý je charakterizován stupněm příslušnosti do této množiny, který leží v intervalu $<0, 1>$, kdy 0 znamená, že objekt není prvkem množiny, 1 značí, že prvek do množiny zcela náleží, a hodnoty mezi 0 a 1 pak udávají míru částečné příslušnosti prvku do fuzzy množiny (Šlechta, 1998, s. 49).

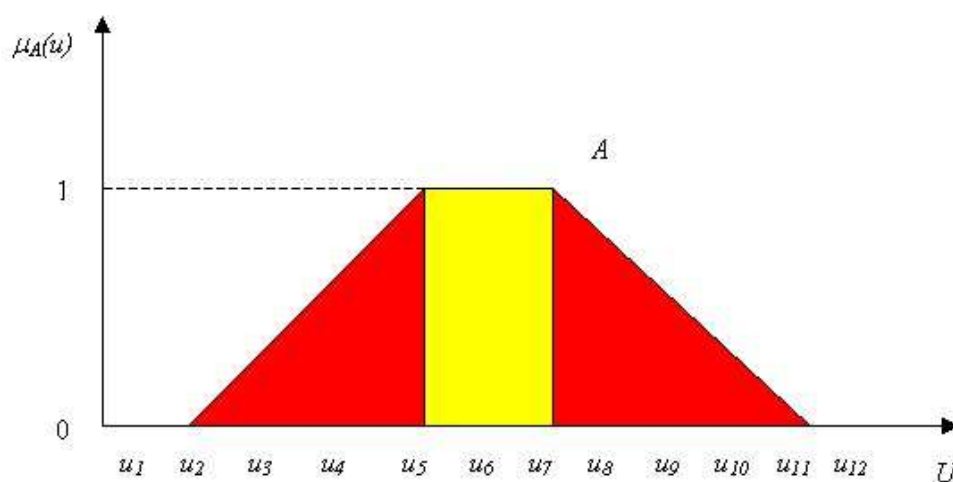
Fuzzy logika má svůj původ na počátku 20. století, kdy britský logik, matematik a filozof *Bertrand Russell* našel v jádru klasické moderní teorie množin a logiky starý a dobře známý logický paradox: *Jistý Kréťan tvrdí, že všichni Kréťané lžou*. Znamená to, že lže i tento Kréťan? Jestliže lže, potom říká pravdu a nelže. Jestliže nelže, potom mluví pravdu, a potom tedy lže. Zde se dostáváme k logickému rozporu (kontradikci), protože tvrzení je jak pravdivé tak i nepravdivé zároveň. Russell našel stejný paradox v teorii množin: *Je množina všech množin, které nejsou členem sebe samých, členem sebe samé?* Jestliže je, potom není; jestliže není, potom je. Tváří v tvář takovýmto paradoxům musí klasická dvouhodnotová logika, pro níž je výrok buďto pravdivý, anebo nepravdivý, kapitulovat, avšak v rámci běžného lidského uvažování tyto výroky de facto žádný paradox nepředstavují a člověku nečiní žádné větší potíže tyto výroky akceptovat, neboť dokáže přijmout to, že „*A a ne-A platí do určité míry zároveň, že mezi vším a ničím neexistuje ostrá dělící čára*“ (Kosko, 1994, s. 6). Takže výše zmiňované paradoxy pak člověk dokáže vnímat např. tak, že jedna polovina výroku je pravdivá a druhá polovina je nepravdivá.

Ve 20. letech 20. století nezávisle na Russellovi polský logik *Jan Lukasiewicz* vypracoval principy **vícehodnotové (multivalentní) logiky**, ve které **tvrzení mohou nabývat částečných pravdivostních hodnot mezi jedničkou (úplnou pravdivostí) a nulou (úplnou nepravdivostí) dvouhodnotové (bivalentní) logiky**, které jsou jen **krajními hodnotami nějakého „pravdivostního kontinua“**. Takto potom můžeme tvrdit, že **něco platí (anebo neplatí) jen do určité míry**. Například dvouleté dítě je zcela určitě mladým člověkem,

devadesátiletý dědeček je zcela určitě starým člověkem, ale pětatřicetiletého muže asi většina lidí bude vnímat jako v něčem ještě mladého a v něčem už jako starého, což není žádný paradox, jestliže do určité míry platí jedno a zároveň i to druhé.

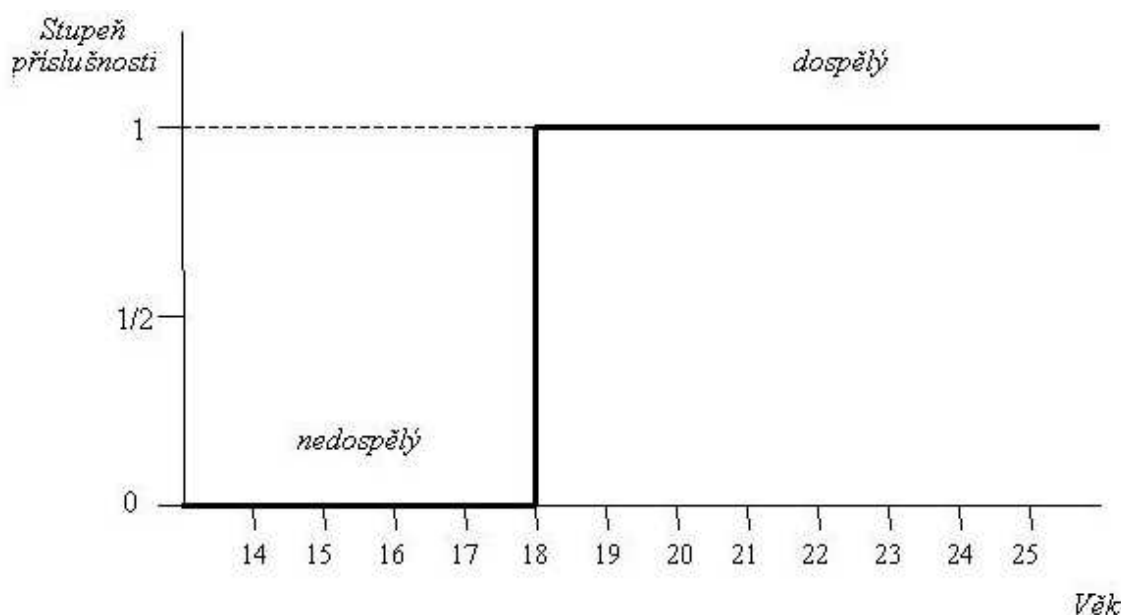
Max Black někdy ve 30. letech 20. století aplikoval multivalentní logiku na klasickou teorii množin (která pracuje na principu náleží/nenáleží), čímž **množiny pozbyly svých jasných kontur a hranic a staly se více rozmazanými, mlhavými, matnými, nejasnými a neurčitými, zkrátka více fuzzy**, takže nyní bylo možné pomocí množin zachytit více než jen dvě krajní pravdivostní hodnoty. Black takto definoval pojem **vágních množin**, čímž položil základy **teorie fuzzy množin**, které se však dostalo větší publicity až po bezmála třiceti letech, kdy matematik Lofti A. Zadeh v roce 1965 publikoval článek s názvem *Fuzzy množiny*.

Fuzzy množina (označme si ji symbolem A) je definována jako množina prvků $u \in U$ (kde U je *univerzem* všech uvažovaných prvků), z nichž každému je přiřazen tzv. **stupeň příslušnosti** $\mu_A(u)$, jehož hodnoty se pohybují v rozmezí $\langle 0, 1 \rangle$ (které lze vyjádřit i v podobě procentuální míry příslušnosti od 0 % do 100 %), přičemž jestliže $\mu_A = 1$, potom daný prvek do množiny A zcela náleží, jestliže $\mu_A = 0$, potom daný prvek do množiny A vůbec nenáleží a jestliže $\mu_A \neq 1$ a zároveň $\mu_A \neq 0$, potom daný prvek do množiny A částečně náleží a částečně nenáleží. Když vyneseme všechny uvažované prvky do grafu, získáme grafické znázornění tzv. **funkce** či **křivky příslušnosti** pro prvky množiny A (Novák, 2000). Na obrázku 90 je vidět, že přechod od příslušnosti „daný prvek určitě nenáleží do množiny A “ (0) k příslušnosti „daný prvek do množiny A určitě náleží“ (1) je pozvolný, čímž vznikají dvě pásma (na obrázku jsou vybarvená červeně), kde příslušnost prvku u do množiny A není jednoznačná, ale je neurčitá, vágní, čili fuzzy.



Obrázek 90: Funkce příslušnosti pro prvky množiny A . Prvky u_1 a u_{12} do množiny A určitě nepatří (bílá barva), prvky mezi u_5 , u_6 a u_7 do množiny určitě patří (žlutá barva) a ostatní prvky patří do množiny A s různým stupněm příslušnosti mezi 0 a 1 (červená barva).

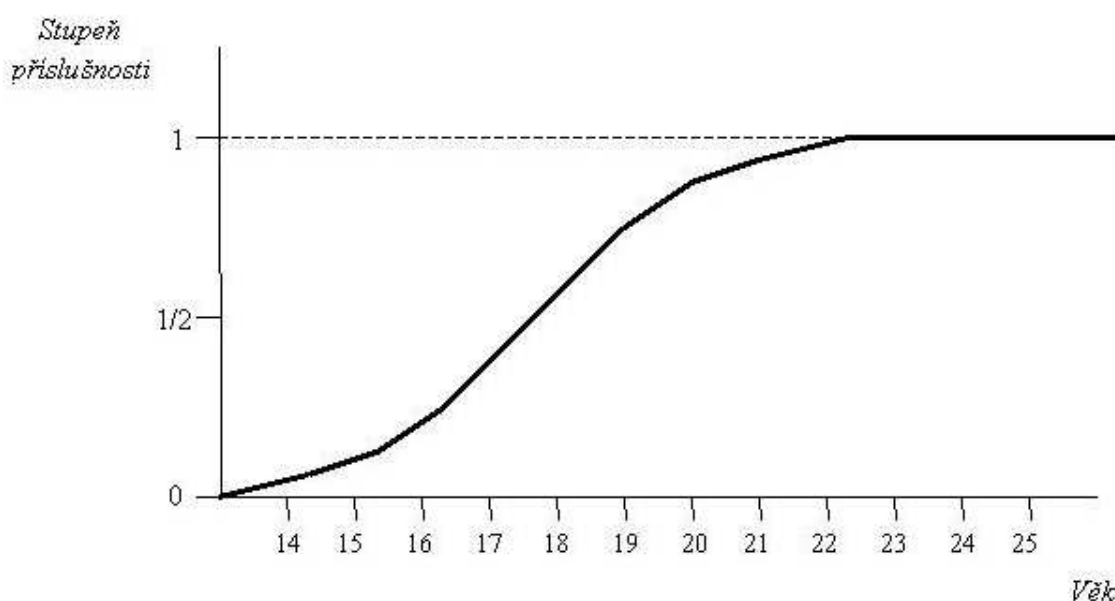
Základní rozdíl mezi **standardními (ostrými) množinami** a **fuzzy (neostrými) množinami**, potažmo mezi **klasickou logikou** a **fuzzy logikou**, se tedy točí okolo toho, co *Aristoteles* nazval **zákonem vyloučeného třetího**. Ve standardní teorii množin, která stojí v základech klasické bivalentní logiky, je pro každý prvek jednoznačně určeno, zda do dané množiny patří (úplné členství v množině), nebo nepatří (žádné členství v množině); žádná třetí možnost neexistuje. Takovou ostrou množinou je např. pojem dospělosti, tak jak vymezen a chápán v zákoně, který stanovuje, že dospělým se člověk stává dovršením 18. roku věku. Křivka příslušnosti jednotlivých roků věku do ostré množiny dospělosti potom bude vypadat tak, jak je to znázorněno na obrázku 91, kde s 18. rokem věku dochází k ostrému přechodu k dospělosti.



Obrázek 91: Stupně příslušnosti jednotlivých roků věku do ostré množiny dospělosti (převzato z Kosko, 1994, s. 35).

Jak zde však již bylo uvedeno, většina pojmů, se kterými lidé ve své mysli operují, není takto přesně definována, ale má naopak nejasné a rozmazané (fuzzy) hranice, takže lidé si potom někdy nejsou tak úplně zcela jistí, jestli mají česnek nebo petržel považovat za zeleninu, neboť se domnívají, že pro zeleninu jsou charakteristické takové vlastnosti jako zelená nebo červená barva, křupavost... a další vlastnosti, jejichž ztělesněním jsou plody jako například rajče, mrkev či salát, které proto lidé považují za mnohem lepší příklad pojmu zelenina než česnek nebo petržel. Jiným příkladem takového vágního pojmu může být *vysoká rychlost*; je otázkou, co ještě patří do množiny *vysoká rychlost* a co už ne: Pokud je rychlost 90 km/h již vysokou rychlostí, proč jí potom není také rychlost 89 km/h? Tento typ paradoxu

je dobře znám již z doby antického Řecka: *Mějme malou hromadu kamení. Pokud přidáme jeden kámen, dostaneme opět malou hromadu. Tedy každá hromada kamení je malá.* Stejný paradox najdeme i u pojmu *dospělost*: Jestliže je 18-letý člověk považován za dospělého, proč za něj nepovažovat člověka, kterému je 17 let a 9 měsíců? Jestliže je možné za dospělého považovat člověka, kterému je necelých 18 let, proč za něj nepovažovat člověka, kterému je necelých 17 let... atd.? Tento typ paradoxu zmizí, jestliže dospělost budeme vidět jako **záležitost míry** a pozvolného přechodu z nedospělosti do dospělosti, jak je to znázorněno na obrázku 92, kde každý jednotlivý rok věku má svůj stupeň příslušnosti do fuzzy množiny dospělosti, který leží někde v intervalu $<0, 1>$.

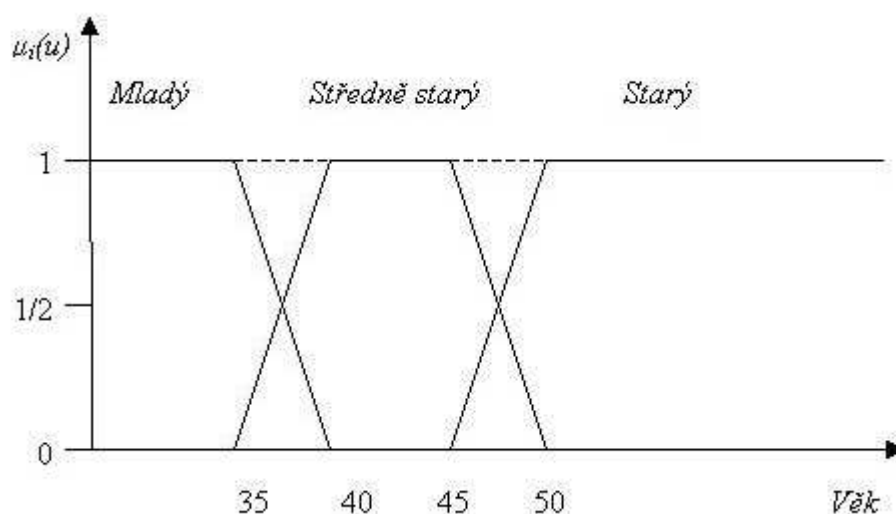


Obrázek 92: Stupně příslušnosti jednotlivých roků věku do fuzzy množiny dospělosti (převzato z Kosko, 1994, s. 35).

Sečteno a podtrženo, fuzzy množiny porušují zákon o vyloučení třetího, neboť jednotlivým prvkům umožňují, aby náležely do více než jedné množiny. V této souvislosti je možné použít terminologii *Kellyho teorie osobních konstruktů* (Mikšík, 2003, s. 185) a fuzzy množiny charakterizovat na základě povahy kontroly, kterou používají vůči svým prvkům: Zatímco klasické ostré množiny představují „zajištěné konstrukty“, které „fixují“ *elementy výlučně do vlastní sféry, [takže] co je dáno do jedné přihrádky, je vyloučené z jakékoli jiné*, fuzzy množiny hrají spíše roli „konstelačních konstruktů“, které „připouštějí, aby *elementy náležely i do jiné oblasti, upravují však jejich členství*“, neboť ve fuzzy logice platí omezení, že stupeň příslušnosti prvku v jednotlivých doplňkových množinách musí dávat dohromady celek, takže jestliže člověk bude ranní vzduch pociťovat jako z 30 % chladný, potom ho musí zároveň

pociťovat jako z 70 % ne-chladný. Z tohoto omezení je patrné, že klasická bivalentní logika, která funguje na principu *člen ze 100 % / člen z 0 %*, představuje pouze speciální případ fuzzy logiky, resp. že fuzzy logika je zobecněním klasické logiky.¹

Kombinací několika navzájem souvisejících fuzzy množin je možné číselně, ale přitom v souladu s neostroší pojmů nadefinovat různé kategorie - tzv. *lingvistické proměnné*, typu *rychlost auta* nebo *věk člověka*. Druhá jmenovaná lingvistická proměnná může zastřešovat například tři fuzzy množiny, které popisují tři různé věkové kategorie: *mladý*, *středně starý* a *starý* - tzv. *atributy* lingvistické proměnné, viz obrázek 93. Tyto lingvistické proměnné mohou



Obrázek 93: Příklad lingvistické proměnné: Věk člověka.

tvorit podmínkovou a důsledkovou část fuzzy inferenčních pravidel typu JESTLIŽE-PAK a tato pravidla – sestávající z jednoduchých nebo složených podmínek a důsledků² – zase mohou být součástí nějakého **fuzzy expertního systému** modelujícího chování určitého

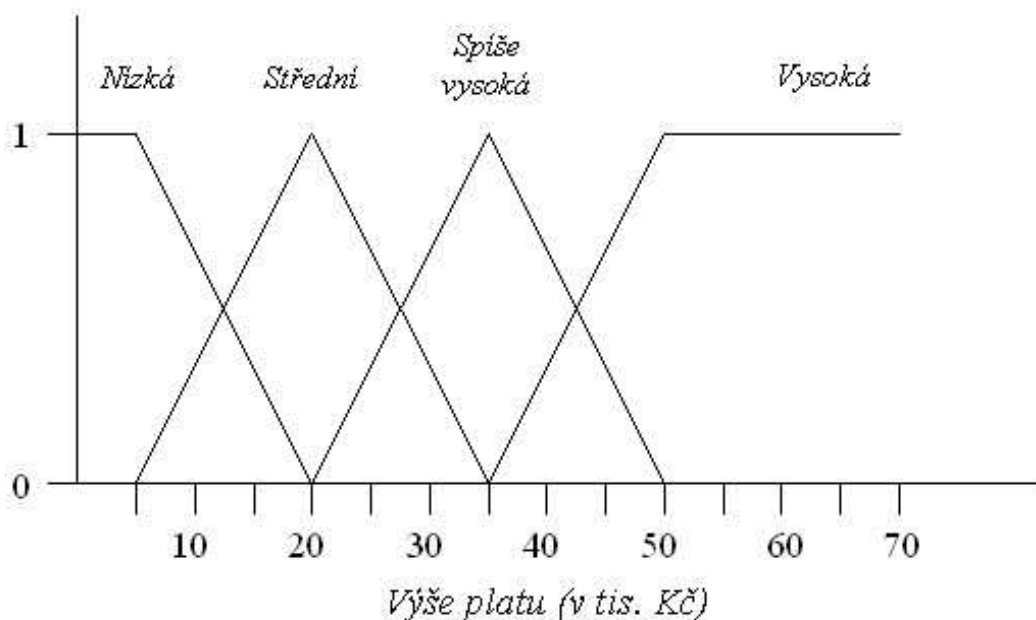
¹ V této souvislosti je potřeba zmínit, že **stupně příslušnosti prvku do fuzzy množiny nelze zaměňovat s procenty teorie pravděpodobnosti**: Zatímco pravděpodobnost vyjadřuje míru toho, že nějaký jev může (ale nemusí) nastat, tj. říká nám něco o tom, zda pozorovaný jev nastane (nebo nenastane) při opakování nějakého pokusu, stupně příslušnosti vyjadřují míru, do jaké se něco vyskytuje. Tento rozdíl mezi teorií pravděpodobnosti a fuzzy logikou je dobře patrný při srovnání následujících dvou tvrzení: *Existuje 30 % pravděpodobnost, že zítra ráno bude chladné počasí.* vs. *Dnešní ráno bylo z 30% studené.* Zatímco první tvrzení sděluje pravděpodobnost toho, že zítra ráno bude chladné počasí, konkrétně, že daný jev (chladné počasí) nastane v průměru ve třech z deseti případů (kdy jsou naměřeny stejné hodnoty pro vlhkost vzduchu... jako na dnešní den), druhé tvrzení konstatuje, že dnešní ráno bylo do určité míry (z 30 %) studené – a zároveň, že bylo právě tak do určité míry (ze 70 %) ne-studené, řekněme např. teplé.

² Složené podmínky/důsledky vznikají spojením dvou a více jednoduchých podmínek/důsledků pomocí logických spojek *a*, *nebo* a *negace*. Interpretace takto spojených fuzzy podmínek/důsledků se liší od interpretace klasických složených podmínek/důsledků, které se skládají z jednoduchých výroků, které mohou být pouze pravdivé (1), nebo nepravdivé (0). Pro interpretaci logických spojek *a*, *nebo* a *negace* v rámci fuzzy složených podmínek/důsledků platí následující vyhodnocovací pravidla: *A*: $\mu_{x \wedge y} = \min(\mu_x, \mu_y)$; *NEBO*: $\mu_{x \vee y} = \max(\mu_x, \mu_y)$; *A*: $\mu_{\neg x} = 1 - \mu_x$ (Fanta, 2001, s. 132-133).

komplexního systému jako je dopravní systém, finanční trhy nebo chování a prožívání člověka za různých situací a podmínek. Takto lze např. vytvořit velice jednoduchý fuzzy expertní systém¹, který napodobuje schopnost personalisty usuzovat na míru motivace zaměstnance v závislosti na výši jeho finanční odměny podle čtyř níže uvedených pravidel:

- 1) **Jestliže** je úroveň finanční odměny nízká, **potom** je míra motivace nízká.
- 2) **Jestliže** je úroveň finanční odměny střední, **potom** je míra motivace střední.
- 3) **Jestliže** je úroveň finanční odměny spíše vysoká, **potom** je míra motivace vysoká.
- 4) **Jestliže** je úroveň finanční odměny vysoká, **potom** je míra motivace střední.

Vstupní proměnnou je zde *výše finanční odměny* a **výstupní proměnnou** pak *míra motivace zaměstnance*. Výše platu nechť je uváděna v tisících Kč a míra motivace nechť je uváděna na desetibodové škále, kde 1 znamená nejnižší míru motivace a 10 nejvyšší míru motivace. Jak je to patrné z výše uvedených pravidel, tyto proměnné sestávají z několika fuzzy (pod)množin: V případě výše finanční odměny to jsou fuzzy množiny *nízká*, *střední*, *spíše vysoká* a *vysoká*; v případě míry motivace to jsou fuzzy množiny *nízká*, *střední* a *vysoká*. Dále je potřeba provést **fuzzifikaci** jednotlivých **vstupních a výstupních hodnot**, která spočívá v tom, že **na základě expertových zkušeností jsou jednotlivým vstupním a výstupním hodnotám přiřazeny určité stupně příslušnosti do jednotlivých fuzzy (pod)množin**, jak je to vidět na obrázcích 94 a 95.



Obrázek 94: Stupně příslušnosti do jednotlivých fuzzy (pod)množin výše platu.

¹ Tento příklad se volně inspirovuje fuzzy expertním systémem, který ve své diplomové práci uvádí Šlechta (1998).

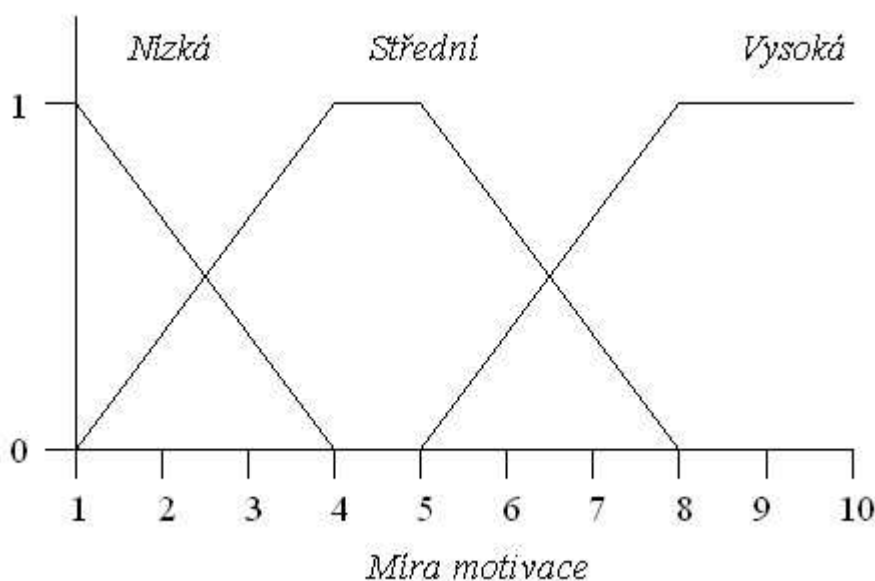
Definujeme-li jednotlivé „platové“ fuzzy množiny výčtem podle obecného vztahu $A = (x_1/\mu_{A1}, x_2/\mu_{A2}, \dots, x_n/\mu_{An})$, potom dostaneme:

$$N_p = \{(5/1), (10/0.7), (15/0.3), (20/0), (25/0), (30/0), (35/0), \dots, (70/0)\}$$

$$S_p = \{(5/0), (10/0.3), (15/0.7), (20/1), (25/0.5), (30/0.3), (35/0), \dots, (70/0)\}$$

$$SV_p = \{(5/0), \dots, (20/0), (25/0.5), (30/0.7), (35/1), (40/0.5), (45/0.3), (50/0), \dots, (70/0)\}$$

$$V_p = \{(5/0), \dots, (35/0), (40/0.5), (45/0.7), (50/1), (55/1), (60/1), (65/1), (70/1)\}$$



Obrázek 95: Stupně příslušnosti do jednotlivých fuzzy (pod)množin míry motivace.

Po definici jednotlivých „motivačních“ fuzzy množin výčtem dostaneme:

$$N_m = \{(1/1), (2/0.7), (3/0.4), (4/0), (5/0), (6/0), (7/0), (8/0), (9/0), (10/0)\}$$

$$S_m = \{(1/0), (2/0.3), (3/0.6), (4/1), (5/1), (6/0.7), (7/0.3), (8/0), (9/0), (10/0)\}$$

$$V_m = \{(1/0), (2/0), (3/0), (4/0), (5/0), (6/0.3), (7/0.7), (8/1), (9/1), (10/1)\}$$

Čtyři výše uvedená (fuzzy) pravidla potom asociují jednotlivé „platové“ fuzzy množiny s příslušnými „motivačními“ fuzzy množinami. To, jaké pravidlo se bude aktivovat, záleží na tom, jaká hodnota finanční odměny se objeví na vstupu, ale vzhledem k tomu, že všechny uvažované vstupní hodnoty byly fuzzifikovány (takže jednotlivé vstupní hodnoty nyní existují v jakési „superpozici“¹ stavů, kdy v různé míře náleží do všech uvažovaných fuzzy

¹ Superpozice je pojem z kvantové mechaniky, kde vyjadřuje tu skutečnost, že každá elementární částice, jako např. foton, představuje vlnu pravděpodobnosti, která je superstavem všech možných stavů, ve kterých se daná elementární částice může s různou pravděpodobností teoreticky nacházet. Akt měření nebo jakýkoli kontakt

(pod)množin), tak se **paralelně a v určité (většinou však nulové) míře aktivují všechna pravidla**. Fuzzy expertní systém takto může obdržet na vstupu informaci, že určitá osoba pobírá za svou práci plat ve výši 20 tisíc Kč; hodnota 20 tisíc Kč patří do určité míry do všech „platových“ fuzzy množin, avšak vzhledem k tomu, že hodnota 20 tisíc Kč náleží do fuzzy množin *nízká*, *spíše vysoká* a *vysoká* z 0 % a do fuzzy množiny *střední* ze 100 %, bude se aktivovat pouze pravidlo číslo 2. Jestliže vstupem bude hodnota 25 tisíc Kč, která z 50 % náleží do fuzzy množiny *střední* a z druhých 50% do fuzzy množiny *spíše vysoká*, budou se aktivovat pravidla číslo 2 a 3, a to každé z 50 %. Platí tedy, že stupně aktivace jednotlivých vstupních fuzzy množin určují stupeň aktivace jednotlivých fuzzy pravidel, resp. výstupních fuzzy množin, které tvoří jejich důsledkovou část.

Abychom získali konkrétnější představu o míře motivace, kterou daná výše finanční odměny vyvolává, je potřeba z parciálně aktivovaných výstupních fuzzy množin udělat jednu jedinou výstupní fuzzy množinu. To lze provést tak, že (1) nejdříve vynásobíme stupně příslušnosti prvků všech výstupních fuzzy množin mírou aktivace odpovídajícího pravidla, resp. jeho důsledkové části¹; (2) potom je potřeba pro každý prvek výstupní proměnné sečíst tyto jejich „zvážené“ stupně příslušnosti do všech výstupních fuzzy množin. Výslednou fuzzy množinu je možné ještě (3) podrobit tzv. **defuzzifikaci**, to když chceme řadu výstupních hodnot s určitým stupněm příslušnosti převést na jednu konkrétní hodnotu: Jeden ze dvou nejpoužívanějších postupů defuzzifikace spočívá v tom, že (3a) za výslednou hodnotu je považován prvek s největším stupněm příslušnosti do výsledné fuzzy množiny; (3b) druhý postup, který se používá u systémů pracujících s numericky definovanými proměnnými, vede k výsledné hodnotě přes výpočet průměru váženého stupni příslušnosti jednotlivých numerických prvků do výsledné fuzzy množiny. (Šlechta, 1998)

Celý postup si můžeme demonstrovat na vstupní hodnotě 25 tisíc Kč, která jednotlivé vstupních fuzzy množiny aktivuje následujícím způsobem:

$$N_p = 0$$

$$S_p = 0.5$$

$$SV_p = 0.5$$

$$V_p = 0$$

s okolním světem způsobuje *kolaps tohoto superstavu do jednoho jediného, většinou toho nejpravděpodobnějšího stavu*, který můžeme „přímo pozorovat“ (či spíše nějak nepřímo zaznamenat) (Gribbin, 1998).

¹ Jinou možnou operací by v této fázi bylo useknutí stupňů příslušnosti jednotlivých prvků všech fuzzy množin na úrovni míry aktivace příslušného pravidla.

Vzhledem k používaným pravidlům se výstupní fuzzy množiny aktivují takto:

$$N_m = 0$$

$$S_m = 0.5$$

$$V_m = 0.5$$

Nyní lze vynásobit stupně příslušnosti prvků množin N_m , S_m a V_m se stupněm aktivace příslušné fuzzy množiny:

$$\begin{aligned} N_m &= \{(1/1), (2/0.7), (3/0.4), (4/0), (5/0), (6/0), (7/0), (8/0), (9/0), (10/0)\} \times 0 \\ &= \{(1/0), (2/0), (3/0), (4/0), (5/0), (6/0), (7/0), (8/0), (9/0), (10/0)\} \end{aligned}$$

$$\begin{aligned} S_m &= \{(1/0), (2/0.3), (3/0.6), (4/1), (5/1), (6/0.7), (7/0.3), (8/0), (9/0), (10/0)\} \times 0.5 \\ &= \{(1/0), (2/0.15), (3/0.3), (4/0.5), (5/0.5), (6/0.35), (7/0.15), (8/0), (9/0), (10/0)\} \end{aligned}$$

$$\begin{aligned} V_m &= \{(1/0), (2/0), (3/0), (4/0), (5/0), (6/0.3), (7/0.7), (8/1), (9/1), (10/1)\} \times 0.5 \\ &= \{(1/0), (2/0), (3/0), (4/0), (5/0), (6/0.15), (7/0.35), (8/0.5), (9/0.5), (10/0.5)\} \end{aligned}$$

Součtem těchto stupňů příslušnosti získáme výslednou fuzzy množinu (VFM):

$$VFM = \{(1/0), (2/0.15), (3/0.3), (4/0.5), (5/0.5), (6/0.5), (7/0.5), (8/0.5), (9/0.5), (10/0.5)\}$$

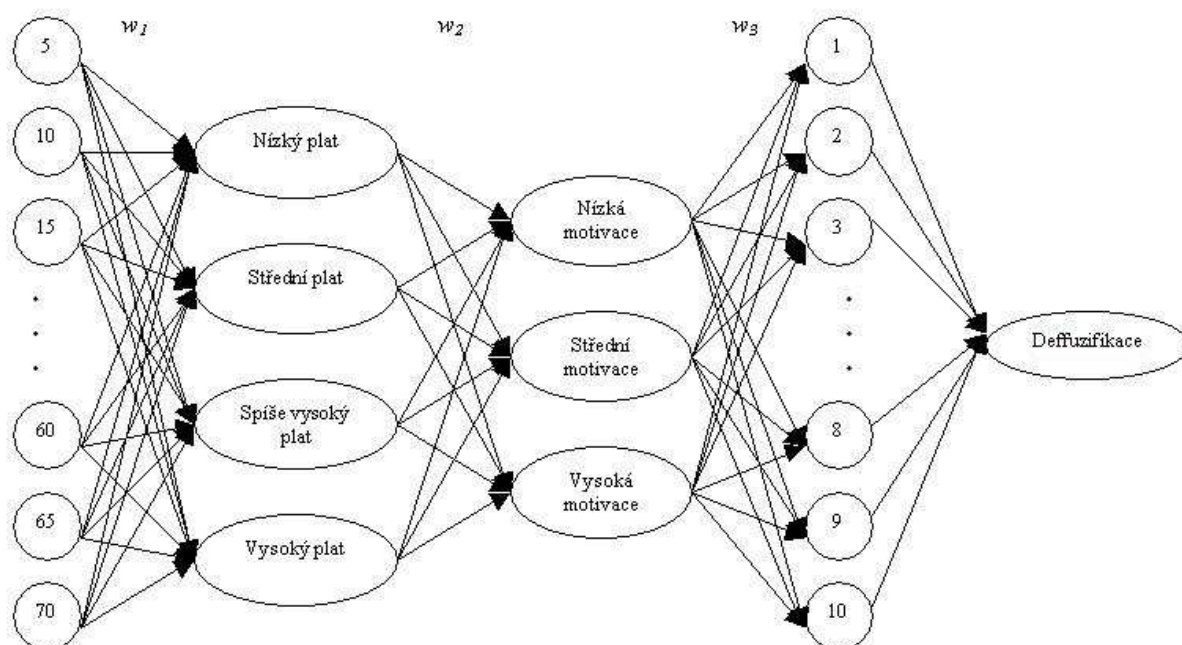
A nakonec množinu VFM defuzzifikujeme pomocí váženého průměru:

$$\frac{1 \times 0 + 2 \times 0.15 + 3 \times 0.3 + 4 \times 0.5 + 5 \times 0.5 + 6 \times 0.5 + 7 \times 0.5 + 8 \times 0.5 + 9 \times 0.5 + 10 \times 0.5}{0 + 0.15 + 0.3 + 0.5 + 0.5 + 0.5 + 0.5 + 0.5 + 0.5 + 0.5} = 6.5$$

Jestliže tedy zaměstnanec bude dostávat plat ve výši 25 tisíc Kč, jeho míra motivace bude mít na desetihodnotové škále hodnotu někde mezi 6 a 7.

Tento jednoduchý fuzzy expertní systém by bylo možné implementovat také v rámci pětivrstevné umělé neuronové sítě (viz obrázek 96): Uzly ve vstupní vrstvě zde reprezentují všechny uvažované vstupní hodnoty výše platu (5-70 tis. Kč). Tyto uzly jsou prostřednictvím vážených spojů propojeny s uzly, které reprezentují čtyři fuzzy (pod)množiny vstupní proměnné: *nízký plat*, *střední plat*, *spíše vysoký plat* a *vysoký plat*. Váha těchto spojů (w_1) odpovídá stupňům příslušnosti jednotlivých vstupních hodnot do všech vstupních fuzzy množin. Vstupní fuzzy množiny jsou dále propojeny se třemi uzly, které představují tři výstupní fuzzy množiny (*nízká motivace*, *střední motivace*, *vysoká motivace*); váhy příslušných spojů (w_2) pak odpovídají asociacím mezi předpokladovými (JESTLIŽE) a

důsledkovými částmi (POTOM) čtyř použitých pravidel. Tyto tři uzly jsou prostřednictvím dalších spojů propojeny s uzly, které odpovídají všem uvažovaným výstupním hodnotám (1-10). Váha spojů mezi těmito uzly (w_3) vyjadřuje stupně příslušnosti těchto výstupních hodnot do jednotlivých výstupních fuzzy množin. Tato výstupní vrstva jednotek je nakonec propojena s jednotkou defuzzifikace, která ji podrobuje operaci váženého zprůměrování.¹ (Šlechta, 1998)



96: Fuzzy expertní systém v podobě pětivrstevné umělé neuronové sítě.

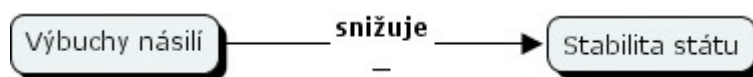
Schopnost fuzzy logiky zachytit nepřesnost a neurčitost prostřednictvím přesného jazyka matematiky je klíčová z hlediska možností matematického modelování komplexních systémů, pro které platí omezení v podobě tzv. Zadehova **principu inkompatibility** – jakési obdoby *Heisenbergova principu neurčitosti*² -, podle kterého s „nárůstem složitosti systému naše schopnost formulovat přesná a přece ještě významná a vypovídající tvrzení o jeho chování klesá, až je dosaženo hranice, za níž se přesnost a významnost (či relevance) stávají prakticky vzájemně se vylučujícími charakteristikami“³ (Zadeh, 1973, s. 28).

¹ Tuto jednoduchou umělou neuronovou síť lze nalézt na příloženém CD v podobě jednoduché excelovské aplikace pod názvem *Fuzzy_Expertní_Systém_Motivace*.

² Podle Heisenbergova principu neurčitosti nelze v mikrosvětě elementárních částic měřit současně polohu a hybnost s naprostou přesností, neboť snaha přesně změřit jednu veličinu vede nevyhnutelně ke zkreslení hodnot druhé veličiny (Greene, 2002).

³ „... as the complexity of a system increases, our ability to make precise and yet significant statements about its behavior diminishes until a threshold is reached beyond which precision and significance (or relevance) become almost mutually exclusive characteristics.“

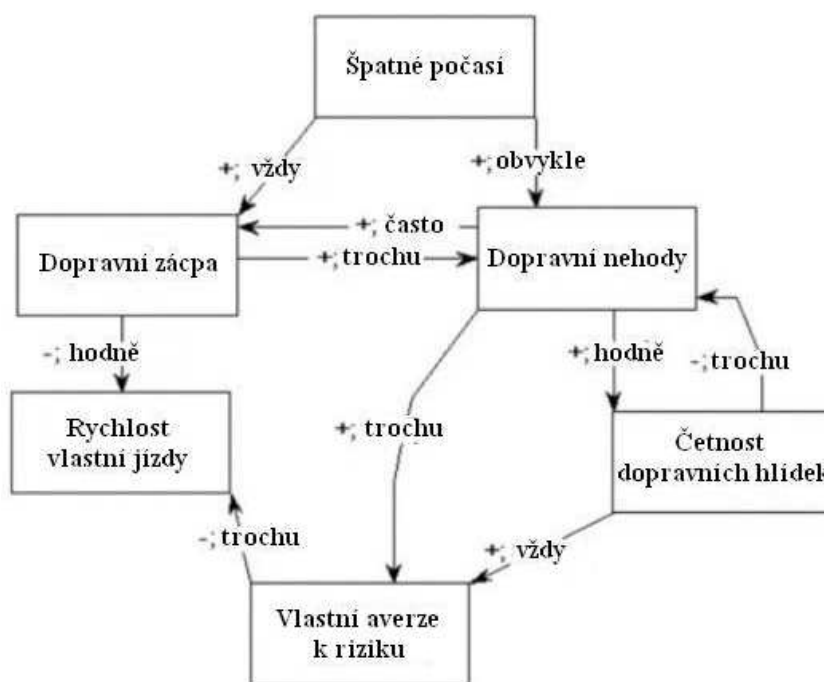
Po tomto stručném obeznámení se se základy fuzzy logiky nyní můžeme přistoupit k samotným fuzzy kognitivním mapám (dále FCM). FCM, se kterými v roce 1986 přišel ve svém článku *Fuzzy cognitive maps* (Fuzzy kognitivní mapy) Bart Kosko, mají podobu **orientovaného grafu, který se skládá z uzlových bodů reprezentujících kvalitativní pojmy a ze synaptických spojů, které vyjadřují kauzální vazby mezi těmito pojmy**. To, čím se fuzzy kognitivní mapy liší od klasických kognitivních map, je povaha pojmových uzlů a kauzálních spojů mezi nimi: V běžných kognitivních mapách jsou pojmové uzly vždy buď aktivované (1), nebo vypnuté (0) a kauzální vztahy mezi těmito pojmovými uzly jsou buď pozitivní, nebo negativní (nebo žádné). **Ve fuzzy kognitivních mapách jsou pojmové uzly fuzzy množinami**, jakými jsou například kategorie *výbuchy násilí* nebo *stabilita státu*: Každá událost do určitého (většinou však nulového) stupně náleží do takové fuzzy množiny, resp. v určité (většinou nulové) míře excituje pojmový uzel, který takovou fuzzy množinu reprezentuje. Pojmový uzel se vždy aktivuje do určitého stupně ($<0, 1>$) a pouze v těch nejjednodušších případech je pojmový uzel buď zapnutý (1), nebo vypnutý (0). Fuzzy pojmový uzel se aktivuje pouze tehdy, když na svém vstupu přijme dostatečné množství kauzálního vlivu z okolních pojmových uzlů, se kterými je propojen synaptickými spoji. V jiném případě se vypne nebo zůstane vypnutý. Když je pojmový uzel aktivní, vysílá určité množství kauzálního vlivu, který se šíří přes synaptická kauzální spojení do dalších pojmových uzlů FCM. Kauzální spoje mezi pojmovými uzly mají v grafu podobu orientované šipky, která reprezentuje fuzzy pravidlo mezi fuzzy množinami: Kauzální pravidlo *Výbuchy násilí snižují stabilitu státu* tak bude mít podobu šipky směřující od pojmového uzlu pro *výbuchy násilí* k pojmovému uzlu pro *stabilitu státu* (obrázek 97). Také **spojení mezi pojmovými uzly mají fuzzy povahu**, neboť umožňují, aby jimi protékalo malé, velké nebo nějaké jiné množství kauzálního vlivu.



Obrázek 97: Fuzzy pravidlo *Výbuchy násilí snižují stabilitu státu*.

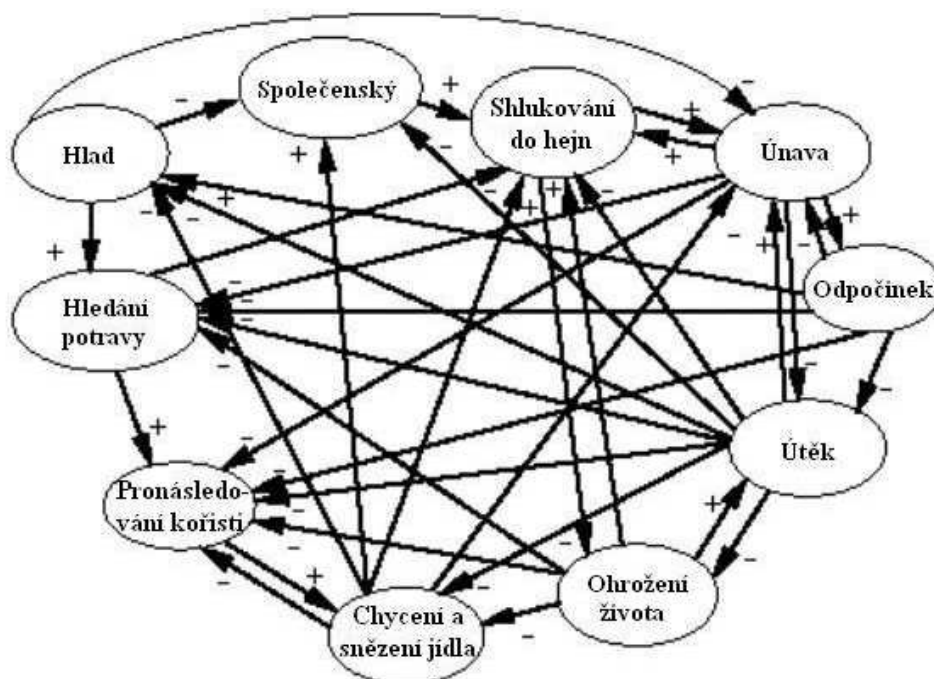
Jako příklad FCM si zde můžeme uvést FCM od Barta Koska (1994, s. 226), na které je zachycena dynamika interakcí mezi počasím a plynulostí dopravy na kalifornských silnicích (viz obrázek 98). V této FCM se nacházejí fuzzy pojmové uzly jako *špatné počasí*, *dopravní zácpa*, *dopravní nehody*, *rychlost vlastní jízdy*, *vlastní odpor k riskování* a *četnost dopravních hlídek*, z nichž každý může nabývat různé míry aktivace ($<0, 1>$), to podle toho, jak moc je

špatné počasí, jak moc vysoká je četnost dopravních hlídek atd. U Fuzzy kauzálních spojů mezi fuzzy pojmovými uzly jsou jejich váhy ($<-1, 1>$) určeny pomocí takových slov jako *trochu*, *hodně*, *vždy*, *často*, *velice*, *poněkud* a *obvykle*. V této FCM se nacházejí rovněž dvě krátké zpětnovazebné smyčky: V prvním případě se jedná o pozitivní zpětnou vazbu: Když dojde k dopravní zácpě, zvýší to poněkud počet dopravních nehod a zvýšený počet dopravních nehod má zase často za následek dopravní zácpu. V druhém případě jde o negativní zpětnou vazbu, kdy zvýšený počet dopravních nehod způsobuje, že na silnicích se z mnohem větší četností začnou objevovat dopravní hlídky, což má za následek mírné snížení počtu dopravních nehod.¹ Podobným způsobem lze modelovat dynamiku chování jakéhokoli komplexního dynamického systému, včetně živých organismů nebo procesu usuzování - takto jsou například na obrázcích 99 a 100 zobrazeny dvě fuzzy kognitivní mapy modelující dynamiku chování „virtuálního“ delfína, resp. proces diferenciální diagnózy specifické jazykové poruchy (*specific language impairment*, SLI).

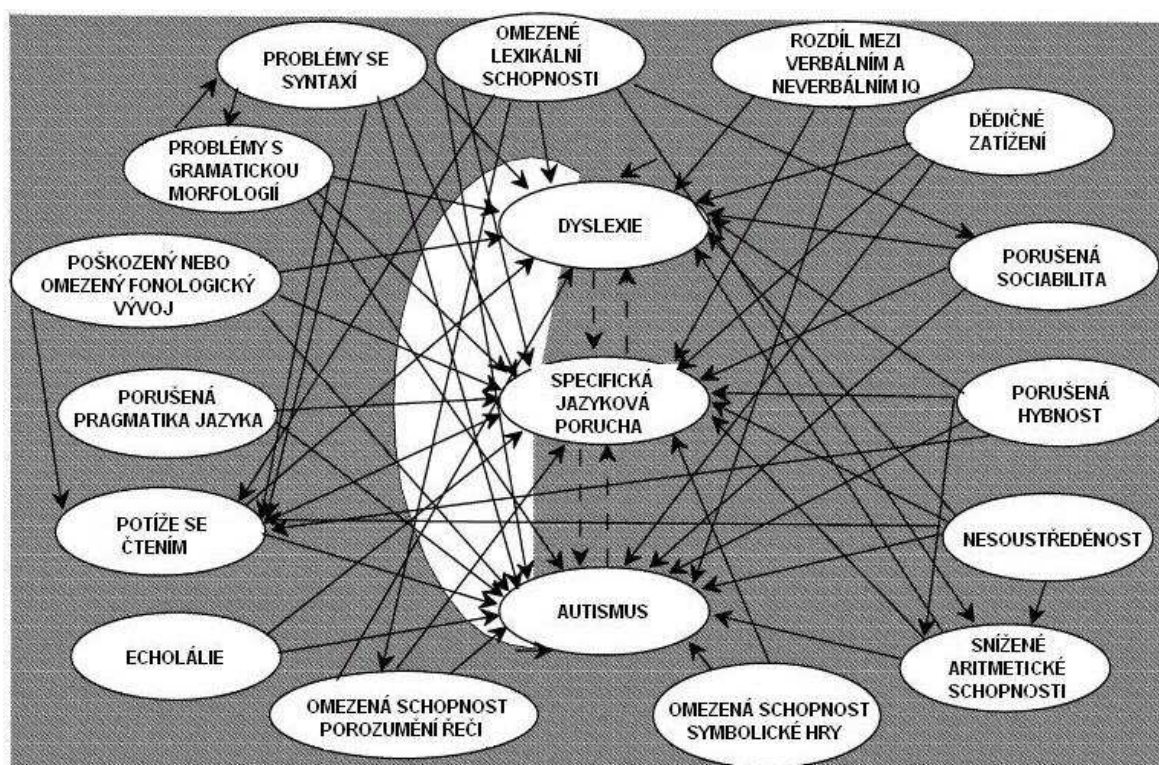


Obrázek 98: FCM dynamiky interakce klimatických podmínek a plynulosti dopravy na kalifornských silnicích (s mírnými úpravami převzato z Kosko, 1994, s. 226).

¹ Na přiloženém CD se pod názvem *FCM_Doprava* nachází excelovská aplikace, která čtenáři umožňuje prozkoumat dynamiku chování této FCM v závislosti na jejím výchozím stavu. V souboru *FCM_Doprava_GA* je pak obsažena stejná FCM, jejíž váhová matice (tj. síla kauzálních spojů mezi jednotlivými pojmovými uzly) však nebyla nastavena nějakým expertem, ale s pomocí optimalizační metody genetických algoritmů (viz 234-239), která byla použita k minimalizaci rozdílu mezi aktuální a požadovanou (cílovou) posloupností stavových vektorů, které byly vygenerovány původní expertní FCM, jejíž váhy byly nastaveny na základě (Koskova) expertního úsudku. Tato druhá FCM tak demonstruje možnosti určitého specifického typu učícího algoritmu, který by mohl sloužit jako alternativa k tradičním učícím algoritmům (typu Hebbova korelačního učícího algoritmu či Hebbova diferenciálního učícího algoritmu, viz 185-187), které se jinak obvykle používají k nastavení váhové matice FCM na základě prezentovaných dat o reálném chování dynamického systému.



Obrázek 99: FCM modelující dynamiku chování „virtuálního“ delfína. (s mírnými úpravami převzato z Dickerson, Kosko, 1996)



Obrázek 100: FCM modelující a podporující proces diferenciální diagnózy specifické jazykové poruchy (specific language impairment, SLI), kterou je v mnoha případech velice obtížné diagnostikovat, neboť symptomy této poruchy se do značné míry překrývají s příznaky dalších poruch jako je těžká dyslexie nebo autismus. Tento výpočetní model byl vytvořen na základě velkého souboru (expertních) poznatků o SLI, které byly shromážděny z veškeré dostupné literatury věnované této poruše. (převzato z Georgopoulos, Malandraki, Stylios, 2003)

K precizaci práce s fuzzy kognitivními mapami je potřeba zavést následující matematický aparát, který je do značné míry totožný s tím, který využívají umělé neuronové sítě (Mls, 2005): Každý fuzzy pojmový uzel FCM C_i je ohodnocen číslem A_i^t , které vyjadřuje míru aktivace daného pojmu. Jestliže je v FCM n pojmů, potom vektor $A^t = [A_1^t, A_2^t, \dots, A_n^t]$ vyjadřuje stav FCM v čase t , přičemž A_i^t udává aktivační úroveň pojmu C_i v čase t . Váhová matice synaptických spojů W je definována jako matice $n \times n$, kde w_{ij} je váha synaptického spoje mezi pojmovými uzly C_i a C_j (ve směru od C_j k C_i).

$$W = \begin{bmatrix} w_{11} & w_{12} & \dots & \dots \\ w_{21} & w_{22} & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & w_{nn} \end{bmatrix}, \text{ kde } w_{ii} = 0$$

Stav FCM, resp. aktivační úroveň všech pojmových uzlů FCM se modifikuje současně (tzv. *synchronní aktualizace*), takže $A^{t+1} = [A_1^{t+1}, A_2^{t+1}, \dots, A_n^{t+1}]$, přičemž A_i^{t+1} , $i = 1, 2, \dots, n$ se počítá podle vztahu:

$$A_i^{t+1} = f\left(\sum_{j=1}^n A_j^t w_{ji}\right)$$

což znamená, že pro každý jednotlivý pojmový uzel se počítá suma násobků aktivace každého pojmu s vahou kauzálního vlivu, který daný pojem spojuje s pojmovým uzlem, jehož stav (tj. míru aktivace) v čase $t+1$ se snažíme vypočítat. Důležitou roli v tomto vztahu má prahová či přechodová funkce f , která zajišťuje, že hodnota aktivace pojmových uzlů nabývá pouze povolených hodnot. Častými prahovými funkcemi jsou:

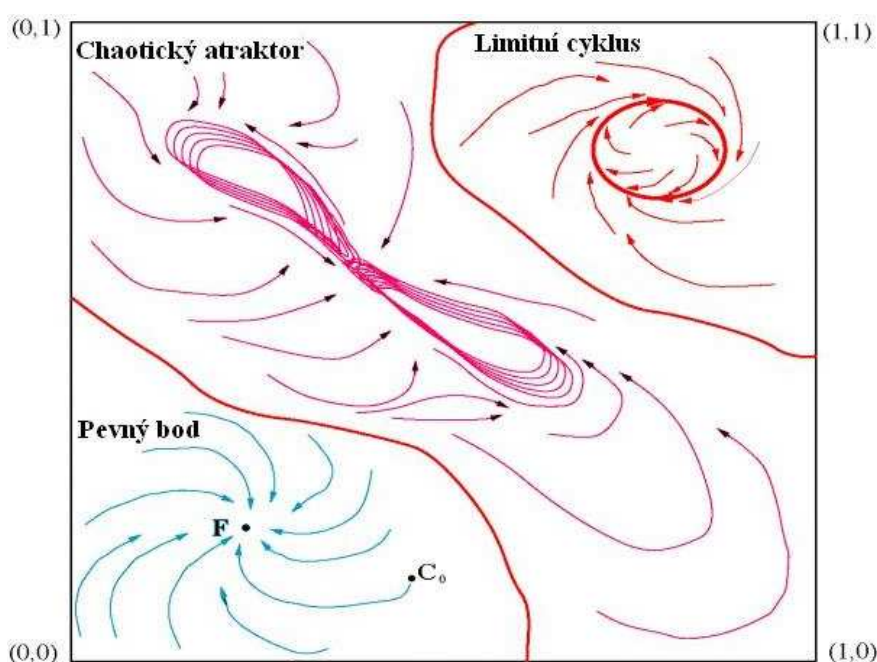
$$f(x) = \begin{cases} x & \text{pro } x \in (0, 1) \\ 1 & \text{pro } x \geq 1 \\ 0 & \text{pro } x \leq 0 \end{cases} \quad f(x) = \begin{cases} 0 & \text{pro } x < 0 \\ 1 & \text{pro } x \geq 0 \end{cases} \quad f(x) = \begin{cases} 1 & \text{pro } x > 0 \\ 0 & \text{pro } x = 0 \\ -1 & \text{pro } x < 0 \end{cases}$$

Samotná práce s FCM při modelování dynamického systému začíná podobně jako při používání příčinného smyčkového diagramu nebo kognitivní mapy, tj. zachycením hypotéz o kauzální struktuře daného dynamického systému v podobě orientovaného grafu. Díky matematickému aparátu spojenému s FCM však lze prostřednictvím FCM mnohem přesněji a důsledněji sledovat a testovat důsledky, které vyplývají z přijatých předpokladů ohledně

příčinných vztahů determinujících chování studovaného dynamického systému: Po sestavení diagramu kauzálních vztahů mezi jednotlivými pojmovými uzly se nastaví výchozí aktivační úroveň každého pojmového uzlu a váhy jednotlivých kauzálních spojů mezi nimi, a pojmovým uzlům se nakonec umožní, aby na sebe navzájem působily, a to dokud

- není dosaženo rovnovážného stavu, kdy se pozitivní a negativní kauzální vlivy mezi pojmovými uzly navzájem vyruší a systém se usadí v nějakém **pevném bodě**, nebo dokud
- není dosaženo nějakého **limitního cyklu**, kdy se systém dostane do smyčky s určitou periodou, takže se bude vždy po stejném počtu kroků vracet do stejného stavu, anebo dokud
- se v systému neprojeví **chaotické chování**.

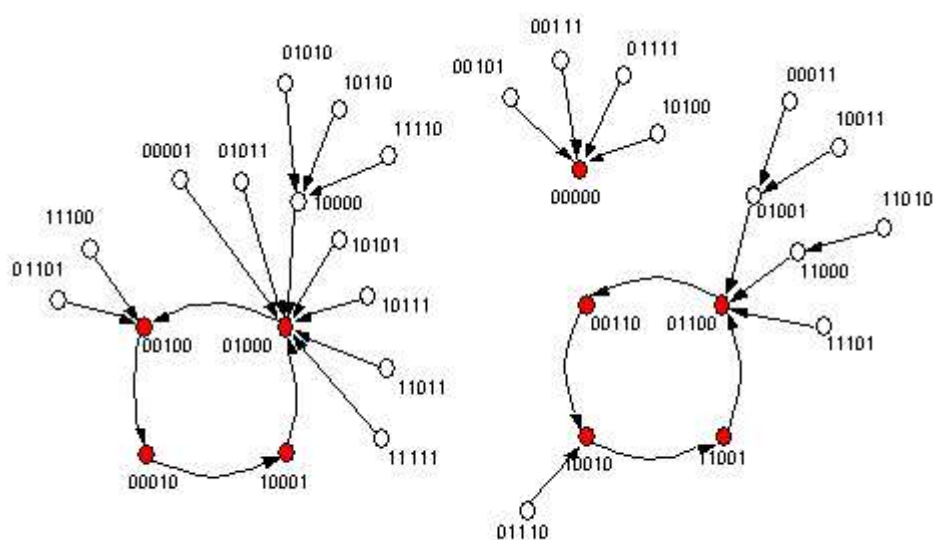
Vzhledem k tomu, že stav FCM je v každém okamžiku dán n -vektorem aktivací všech jejích n pojmových uzlů, lze její chování (a potažmo tedy také chování modelovaného systému) vizualizovat prostřednictvím fázového diagramu jako trajektorii vykreslenou historií stavových vektorů FCM v n -dimenzionálním **stavovém prostoru**, kde n je počet pojmových uzlů tvořících FCM – na obrázku 101 jsou takto schematicky znázorněny tři základní mody



Obrázek 101: Dvourozměrný stavový prostor FCM sestávající ze dvou pojmových uzlů. Na obrázku jsou schematicky zachyceny tři základní způsoby dlouhodobého chování takové FCM v podobě pevného bodu (kdy se stavový vektor FCM nemění), limitního cyklu (kdy stavový vektor FCM opakuje konečnou posloupnost stavů) a chaotického atraktoru (kdy se stavový vektor FCM neperiodicky mění). (s mírnými úpravami převzato z Dickerson, Kosko, 1996)

chování dynamického systému pro FCM sestávající ze dvou pojmových uzlů. Chování FCM, resp. strukturu jejího stavového prostoru lze stejně názorně zachytit také prostřednictvím diagramu stavových přechodů – na obrázku 102 je tímto způsobem znázorněno chování jednoduché binární FCM sestávající z pěti pojmových uzlů (jejichž míra aktivace může nabývat pouze dvou možných hodnot, 0/1, proto binární FCM); z diagramu je na první pohled zřejmé, že chování této FCM (a jí modelovaného dynamického systému) charakterizují tři limitní množiny (na obrázku zvýrazněné červenou barvou) - jeden pevný bod a dva limitní cykly.

Z hlediska uživatelské vstřícnosti je důležité, že při používání FCM k modelování komplexních dynamických systémů se pracuje většinou s pojmy běžného a dobře srozumitelného jazyka a nikoli se složitými diferenciálními rovnicemi. K uživatelské vstřícnosti FCM přispívá také to, že chování každé navržené FCM lze poměrně snadno modifikovat - tím, že se do ní přidají nové pojmové uzly nebo nové synaptické spoje, tím že se v polaritě (+/-) nebo v intenzitě (<-1, 1>) upraví stávající nastavení kauzálních spoju nebo



Obrázek 102: Diagram stavových přechodů jednoduché binární FCM sestávající z pěti pojmových uzlů, které mohou nabývat pouze dvou možných hodnot (0/1). V diagramu jsou přehledným způsobem zachyceny přechody mezi jednotlivými 32 (2^5) stavovými vektory této FCM. Na diagram se lze také dívat jako na reprezentaci struktury pětidimenzionálního stavového prostoru FCM, ve kterém se (vzhledem k povaze použité přechodové funkce) nachází 32 identifikovatelných pozic. Na první pohled jsou zde dobře patrné tři limitní množiny (červené tečky) – jeden pevný bod a dva limitní cykly. (s úpravami převzato z MIs, 2005)

tím, že se jinak nastaví výchozí hodnoty aktivace jednotlivých pojmových uzlů. Tyto úpravy většinou provádí expert, který je detailně obeznámen se zákonitostmi daného dynamického

systemu, použít však lze také automatický učicí algoritmus, který na základě dat o reálném chování modelovaného systému dokáže vygenerovat tu správnou váhovou matici synaptických spojů mezi jednotlivými pojmovými uzly, která odráží skutečnou kauzální strukturu modelovaného systému. Při použití takového učicího algoritmu se nejdříve data o reálném chování modelovaného systému převedou do podoby posloupnosti stavových vektorů FCM, tj. do podoby informací o míře, v jaké jsou jednotlivé pojmové uzly aktivní v sérii diskrétních časových kroků sledovaného období. Tyto informace pak slouží jako vstup do samotného učicího algoritmu. Existuje přitom několik různých učících algoritmů. Mezi ty nejznámější patří **Hebbův korelační učicí algoritmus** (dále HKU) a **Hebbův diferenciální učicí algoritmus** (dále HDU). HKU vychází z předpokladu, že jsou-li dva pojmové uzly současně aktivní, je zde vysoká pravděpodobnost, že mezi těmito pojmovými uzly existuje také nějaká příčinná souvislost. V případě HKU je tak vznik a nastavení váhy kauzálního spoje založeno na korelaci aktivity dvou pojmových uzlů; samotná změna váhy kauzálního spoje mezi pojmovými uzly se pak počítá podle vztahu:

$$\dot{e}_{ij} = -e_{ij} + C_i(x_i)C_j(x_j)$$

kde $-e_{ij}$ je aktuální váha kauzálního spoje mezi pojmovými uzly C_i a C_j a $C_i(x_i)$ a $C_j(x_j)$ představují míru aktivace pojmových uzlů C_i a C_j v daném časovém okamžiku. S použitím HKU pro nastavení váhové matice FCM je spojena jedna zásadní nevýhoda – HKU často nachází velké množství falešných kauzálních vztahů tam, kde ve skutečnosti žádné nejsou. Tuto slabinu a omezení HKU se Kosko pokusil překonat s pomocí HDU, který je založen na úsudkové verzi gestaltistického zákona „společného osudu“, podle kterého máme při organizaci našeho vjemového pole tendenci si spojovat předměty se společným pohybem (směrem, rychlostí). To, co platí, pro naše vnímání, platí do značné míry i pro naše usuzování na kauzální souvislosti mezi jevy. Britský logik a filozof *John Stuart Mill* v této souvislosti hovořil o tzv. metodě **průvodní variace** (*concomitant variation*), která je jedním z nástrojů induktivního usuzování umožňujících kauzální analýzu, tedy zjišťování příčinných vztahů (Ferjenčík, 2000, s. 26-29). Podstata této metody spočívá ve variování intenzity té okolnosti, o níž předpokládáme, že je příčinou studovaného jevu (za souběžného udržování ostatních okolností v konstantní a neměnné podobě). V případě, že se studovaný jev bude měnit současně se změnami intenzity dané okolnosti, potom lze s poměrně vysokou mírou jistoty tvrdit, že mezi danou okolností a jevem existuje kauzální spojení. Jinak řečeno, máme tendenci si kauzálně spojovat dva jevy tehdy, jestliže se tyto dva jevy mění společně. Při hledání kauzálních souvislostí tak HDU na rozdíl od HKU mezi sebou nenásobí míru aktivace



Obrázek 103: Gestaltistický zákon „společného osudu“, který způsobuje, že máme tendenci si spojovat objekty se společným pohybem (směrem či rychlostí).

dvou pojmových uzlů, ale to, jak se míra aktivace těchto dvou pojmových uzlů mění, tedy jejich derivace¹:

$$\dot{e}_{ij} = -e_{ij} + \dot{C}_i(x_i) \dot{C}_j(x_j)$$

Kosko (1994, s. 234) princip HDU srozumitelně popisuje následujícím způsobem: „*Mějme pojmové uzly A a B. [...] Předpokládejme, že se [míra aktivace] A mírně zvýší. Potom je rozdíl mezi novým A a starým A pozitivní. Takže změna A je pozitivní (+). Řekněme, že se nyní mírně zvýší také [míra aktivace] B. Takže také změna B je pozitivní (+). A plus krát plus se rovná plus (+1 x +1 = +1). Přítel mého přítele je můj přítel. Takže k hraně vedoucí od A k B přičtete malé číslo. Řekněme, že nyní [míra aktivace] A klesne. Změna je potom negativní (-). Řekněme, že klesne také [míra aktivace] B. Takže také změna B je negativní (-). A jestliže nyní vynásobíte tyto dvě změny, dostanete pozitivní hodnotu (-1 x -1 = 1). Nepřítel mého nepřítel je můj přítel. Takže opět přičtete malé číslo ke hraně [mezi A a B]. Jestliže se tedy A a B mění společně, potom odhadujeme, že mezi nimi existuje kauzální spojení. Opak nastane, jestliže se budou pohybovat opačným směrem. Řekněme, že [míra aktivace] A se zvýší a [míra aktivace] B klesne. Vynásobením pozitivní a negativní hodnoty pak dostaneme negativní hodnotu (+1 x -1 = -1). Přítel mého nepřítel je můj nepřítel. To samé nastane, jestliže [míra aktivace] A klesne a [míra aktivace] B stoupne. Nepřítel mého přítele je můj nepřítel. V obou dvou případech odečtete od hrany [mezi A a B] malé číslo a ta má tendenci se měnit v negativní*

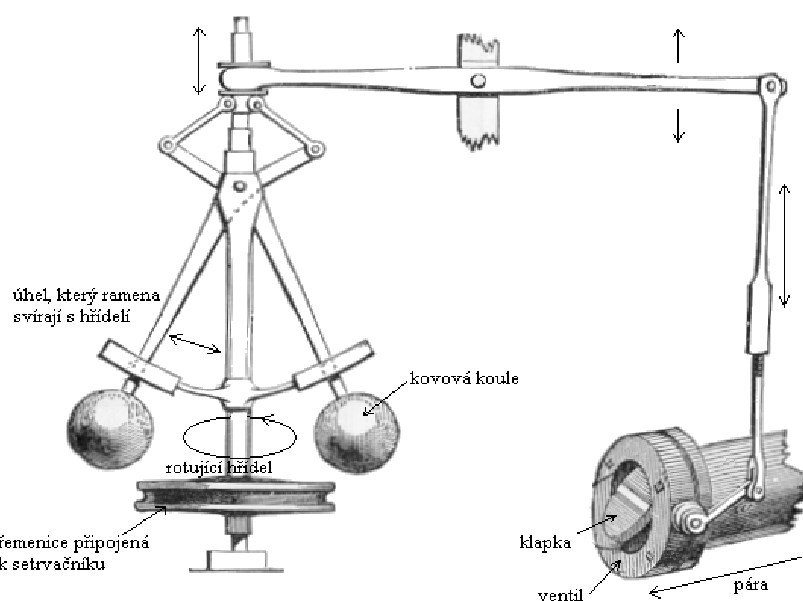
¹ Viz poznámka pod čarou 1 na straně 154.

[inhibiční] spojení. Jestliže se ani A, ani B nemění, potom dostanete nulu a nulu přičtete k hraně [mezi A a B].“¹

Rovněž HDU má určité slabiny, které tomuto učicímu algoritmu brání objevovat určitý druh kauzálních souvislostí mezi jevy. Proto se objevila řada návrhů, jak základní algoritmus HDU pozměnit a jeho „krátkozrakost“ tak korigovat (viz Huerga, 2005).

2.2.1.4 Dynamická kognitivní věda

Inspirace kognitivní vědy myšlenkami teorie dynamických systémů vedla někdy okolo poloviny 90. let 20. století ke vzniku **dynamické kognitivní vědy** (*dynamic cognitive science*), která na rozdíl od tradiční, **výpočetní kognitivní vědy** nepovažuje kognitivní systém za systém výpočetní, ale především za systém dynamický. Kognitivní vědci *Tim van Gelder* a *Robert F. Port* (1995) ilustrují rozdíl mezi přístupem výpočetní a dynamické kognitivní vědy na odlišných popisech fungování **Wattova odstředivého regulátoru** (viz obrázek 104) - paradigmatického příkladu dynamického systému, který se často objevuje v nejrůznějších

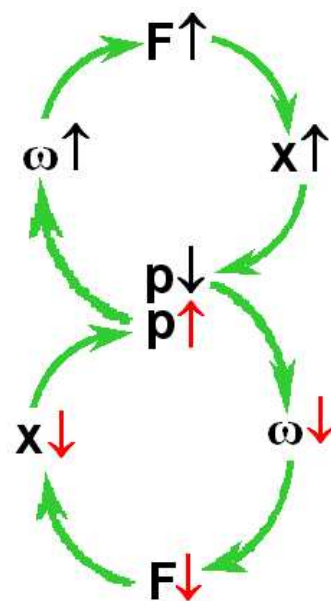
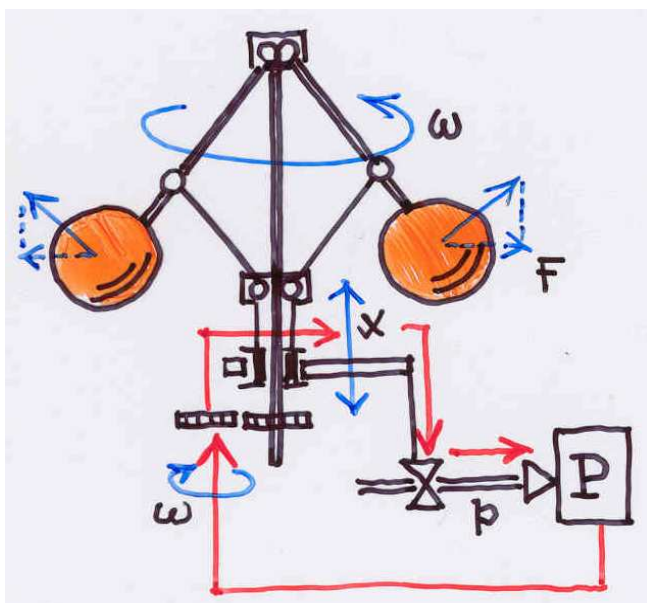


Obrázek 104: Wattův odstředivý regulátor.

¹ „Let A and B be FCM nodes. [...] Suppose A goes up a little. Then the difference between the new A and the old A is positive. So the change of A is positive (+). Now say B goes up a little too. So the change of B is positive (+) too. Then plus times plus equals plus ($+1 \times +1 = +1$). The friend of my friend is my friend. So you add a little number to the edge from A to B. Now say A falls. Then the change is negative (-). Say B falls too. So the B change is negative (-) too. Then if you multiply the two changes you get a positive value ($-1 \times -1 = +1$). The enemy of my enemy is my friend. So again you add a small number to the edge. So if A and B change together, we guess a causal link between them. The reverse happens if they move in opposite directions. Say A goes up and then B goes down. Then multiply a positive by a negative to get a negative ($+1 \times -1 = -1$). The friend of my enemy is my enemy. The same happens if A goes down and B goes up. In both cases you subtract a small number from the edge and it tends to move to a negative link. If neither A nor B changes, you get zero and add zero to the edge.“

úvodech do teorie dynamických systémů – a **hypotetického výpočetního regulátoru**. Wattův odstředivý regulátor představuje jednoduché a elegantní řešení jednoho z nejpálčivějších problémů průmyslu 18. století. Ke konci 18. století se v průmyslové výrobě začal hromadně využívat parní stroj, který dokázal transformovat tepelnou energii vodní páry na mechanickou práci: Nejdříve se v parním kotli s pomocí tepla (uvolněného při spalování nějakého paliva) a vody vyrobila vodní pára s požadovanou teplotou a tlakem; takto vyrobená vodní pára pak tlakovou silou působila na píst a posunovala ho ve válci; tento přímočarý pohyb pístu se pak převedl na otáčivý pohyb setrvačníku, který prostřednictvím systému převodů poháněl nejrůznější průmyslové stroje a zařízení. K tomu, aby tyto stroje správně fungovaly, bylo nutné, aby jejich chod byl plynulý, tj. aby byly poháněny stálou a neměnnou rychlostí. Problém však byl v tom, že rychlost pohybu setrvačníku často kolísala v závislosti na řadě různých faktorů, například v závislosti na změnách teploty v topeništi parního kotle či v závislosti na aktuálním pracovním zatížení parního stroje (daném počtem poháněných strojů). Technici tak stáli před problémem, jak zajistit konstantní rychlost setrvačníku navzdory neustálému kolísání teploty a tlaku vodní páry či pracovního zatížení parního stroje.

S geniálně jednoduchým řešením tohoto problému přišel v roce 1784 skotský fyzik a vynálezce *James Watt*, který vynalezl to, co se dnes nazývá *Wattův odstředivý regulátor*: K setrvačníku parního stroje je připevněna hřídel, na kterou jsou zavěšena dvě sklopná ramena; na konci každého ramene je kovová koule; ramena jsou připojena ke klapkovému ventilu a to takovým způsobem, že čím více se ramena zvednou (tj. čím větší úhel budou ramena svírat s hřídelí), tím menší množství páry projde klapkovým ventilem. Regulace rychlosti setrvačníku pak probíhá následujícím způsobem: Jak se hřídel otáčí okolo své osy, odstředivé síly způsobují, že se ramena zvedají a zvedají se tím více, čím se hřídel otáčí rychleji. Zvednutí ramen však způsobuje to, že se klapkový ventil uzavírá, čímž se omezuje přívod vodní páry. To má za následek zpomalení setrvačníku, resp. hřídele, která je k němu připojena. Zpomalení setrvačníku pak vede ke snížení odstředivé síly a k poklesnutí ramen, což zase způsobuje pootevření klapkového ventilu a zvýšení přívodu vodní páry... (viz obrázek 105). Při správném nastavení takto Wattův odstředivý regulátor dokáže udržet konstantní rychlost setrvačníku a to navzdory neustálému kolísání řady různých faktorů ovlivňujících jeho rychlost.



Obrázek 105: Znáznornění činnosti Wattova odstředivého regulátoru v podobě schematického nákresu a příčinného smyčkového diagramu, ze kterého je na první pohled zřejmá pro komplexní dynamické systémy tak charakteristická cirkulární kauzalita, spočívající ve vzájemném zpětnovazebním ovlivňování jednotlivých prvků systému (ω – rychlost otáčení setrvačníku, resp. hřídele, F – odstředivá síla působící na kovové koule na konci dvou ramen, x – míra posunutí táhla, p – množství páry uvádějící do pohybu píst, resp. setrvačnick). (Schematický nákres Wattova regulátoru převzat z ppt prezentace Ivana M. Havla k přednáškám „Přirozené a umělé myšlení“ konaným ve školním roce 2005/2006 na pracovišti Centra Teoretických Studií v Praze)

Chování Wattova odstředivého regulátoru lze přesně popsat a zachytit s pomocí několika málo parametrů, jedné numerické proměnné a prostřednictvím matematické rovnice, která definuje chování této proměnné v čase. Klíčovou proměnnou je zde úhel, který svírají ramena s hřídelí (θ); chování této proměnné popisuje následující diferenciální rovnice:

$$\frac{d^2\theta}{dt^2} = (n\omega)^2 \cos\theta \sin\theta - \frac{g}{l} \sin\theta - r \frac{d\theta}{dt}$$

kde n je převodová konstanta, ω je rychlost otáčení hřídele, g je gravitační konstanta, l je délka ramen, a r je konstanta míry tření na čepech. Tato diferenciální rovnice nám říká, že míra změny úhlu mezi rameny a hřídelí ($d^2\theta/dt^2$, tj. druhá derivace¹ úhlu θ podle času t) je funkcí momentálního úhlu mezi rameny a hřídelí (označeného stavovou proměnnou θ), rychlosti, s jakou se tento úhel právě mění ($d\theta/dt$, tj. derivace úhlu θ podle času t), a momentální rychlosti otáčení hřídele (ω). Jinak řečeno, rovnice zachycuje to, jak se mění samotná změna úhlu svíraného rameny a hřídelí v závislosti na současném úhlu, na způsobu,

¹ Tzn. derivace derivace (viz také poznámka pod čarou 1 na straně 154).

jakým se tento úhel právě mění, a na rychlosti otáčení hřídele. Jediné, co se mění v systému definovaném touto diferenciální rovnicí, je úhel mezi rameny a hřídelí θ (a jeho derivace); vše ostatní jsou (řídící) parametry (ω , n , g , l , a r), o nichž se předpokládá, že jejich hodnota je stabilní a v čase se nijak nemění. Nastavení hodnot těchto parametrů pak determinuje konkrétní podobu (tvar) chování proměnné θ v čase.

Aby regulátor mohl vykonávat svoji funkci, musí být – prostřednictvím setrvačníku a hřídele – propojen s parním strojem, který lze rovněž s pomocí několika různých proměnných, parametrů a matematických rovnic popsat jako dynamický systém. Tyto dva systémy (regulátor a parní stroj) jsou přitom propojeny určitým specifickým způsobem, který spočívá v tom, že proměnná v jednom systému (například úhel θ u Wattova regulátoru či rychlost otáčení setrvačníku ω u parního stroje) hraje v druhém systému roli řídícího parametru. Systém několika rovnic takto dokáže popsat komplexní interakci dvou vzájemně se ovlivňujících systémů, které společně dohromady tvoří jeden propojený systém, jehož chování má tendenci být přitahováno ke stabilnímu jednobodovému atraktoru, kterým je požadovaná stabilní rychlost otáčení hřídele.

van Gelder a Port dávají tento Wattův odstředivý regulátor, jehož chování lze úspěšně porozumět prostřednictvím pojmů teorie dynamických systémů, do ostrého kontrastu s hypotetickým výpočetním regulátorem, jehož chování je možné zase naopak úspěšně vysvětlit prostřednictvím pojmových nástrojů výpočetní teorie. Při konstrukci takového výpočetního regulátoru lze vyjít z obecného popisu zpětnovazebních řídicích systémů, které k regulaci a řízení vybraného parametru (v tomto případě rychlosti setrvačníku parního stroje) využívají tři základní druhy informací (Hillis, 2003, s. 121):

- 1) *Jaký je požadovaný stav (cíl)?*
- 2) *Jaký je rozdíl mezi současným a požadovaným stavem (chyba)?*
- 3) *Jaké činnosti zmenší rozdíl mezi současným a požadovaným stavem (reakce)?*

Zpětnovazební řídicí systém tedy dosahuje svého cíle tím, že upravuje a volí svoji reakci v závislosti na druhu detekované chyby. Konkrétní algoritmus výpočetního regulátoru by pak mohl vypadat následujícím způsobem:

1. *Změř rychlost otáčení setrvačníku.*
2. *Porovnej aktuální rychlost otáčení setrvačníku s požadovanou rychlostí.*
3. *Jestliže mezi těmito dvěma rychlostmi není žádný rozdíl, potom se vrať ke kroku 1; v jiném případě*
 - a) *Změř aktuální tlak páry.*
 - b) *Vypočítej požadovanou změnu tlaku páry.*

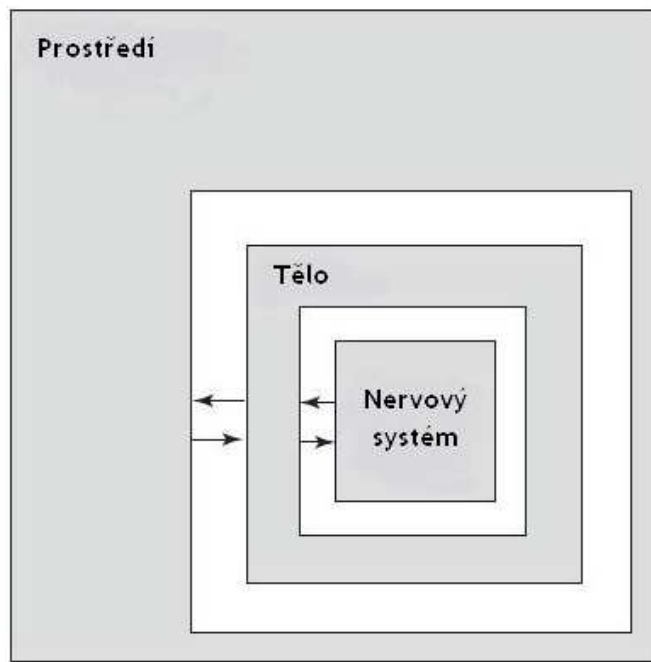
c) *Vypočítej potřebnou úpravu klapkového ventilu.*

4. *Proved' úpravu klapkového ventilu.*

5. *Vrat' se ke kroku 1.*

Výpočetní regulátor pracující na základě takového algoritmu by nejdříve změřil aktuální rychlost otáčení setrvačníku, poté by vytvořil vnitřní symbolické reprezentace tohoto a dalších údajů, následně by na ně aplikoval různé výpočetních operace, které by mu umožnily tyto symbolické reprezentace mezi sebou porovnat a na základě různých transformačních a manipulačních pravidel je změnit a přetransformovat do podoby, která bude odpovědí na otázku „*Jak upravit klapkový ventil, aby se dosáhlo požadované rychlosti otáčení setrvačníku?*“ Poté, co se regulátor k této odpovědi propočítá, převede ji ve fyzickou činnost a změnu nastavení klapkového ventilu do vypočítané polohy. Činnost výpočetního regulátoru takto tvoří neustále dokola se opakující cyklus měření-počítání-konání. Celý systém pak lze charakterizovat jako systém, který je cyklický, sekvenční, reprezentační a výpočetní. Žádnou z těchto charakteristik však nelze aplikovat na Wattův odstředivý regulátor, u kterého není schopnost řízení a regulace rychlosti setrvačníku výsledkem realizace určité posloupnosti výpočetních operací - ve smyslu vytváření, manipulace a transformace vnitřních symbolických reprezentací vnějšího světa -, ale emergentním důsledkem interakce a vzájemné kodeterminace dvou systémových komponent, kde úhel mezi rameny a hřídelí průběžně a v reálném čase moduluje rychlost setrvačníku, která stejným způsobem zase moduluje úhel, který ramena svírají s hřídelí.

K modelování tohoto druhu interakce, kdy dochází ke vzájemné kodeterminaci chování většího množství různých faktorů, se zdá být ideální právě jazyk teorie dynamických systémů (v podobě fázových prostorů, atraktorů, repelorů, basins of attraction, řídicích parametrů, diferenciálních rovnic atp.). Ta se tak stává vhodným výzkumným nástrojem pro představitele teorie **vtělené, situované a distribuované kognice** (*embodied, situated, distributed cognition*), podle nichž **zdrojem inteligentního chování nejsou výhradně výpočetní procesy odehrávající se na vnitřních symbolických reprezentacích, ale především celková interakce organismu se svým prostředím**, v níž hrají výpočetní procesy pouze jednu dílčí - jakkoli důležitou a nepostradatelnou – roli (viz obrázek 106). Tento směr výzkumu dal v polovině 90. let 20. století vzniknout tzv. dynamické kognitivní vědě (Beer, 2000; van Gelder, Port, 1995) - novému pojetí kognitivní vědy, podle kterého kognitivní systémy nejsou (pouze) výpočetní systémy (jak tvrdí tradiční výpočetní kognitivní věda), ale (především) dynamické systémy, jejichž chování je determinováno interakcí většího množství různých faktorů nacházejících se uvnitř i vně organismu.



Obrázek 106: V pojetí dynamické kognitivní vědy je kognice výsledkem interakce mezi nervovým systémem, tělem a vnějším prostředím. (s mírnými úpravami převzato z Beer, 2000)

Příkladem využití teorie dynamických systémů k popisu a explanaci jistého (vývojově) kognitivního jevu je dynamický **model plánovacího pole** (*planning field model*), který je výsledkem snahy vývojové psycholožky *Esther Thelen* a kol. (2001) vysvětlit rozporuplná experimentální data nashromážděná v rámci výzkumu jednoho specifického kognitivního omylu, kterého se dopouštějí děti staré 8-10 měsíců. Jedná se o tzv. *chybu A-ne-B* (*A-not-B error*), kterou objevil švýcarský psycholog *Jean Piaget* při zkoumání vývoje *konceptu stálosti objektu* u malých dětí. Ke zkoumání vývoje tohoto pojmu Piaget vymyslel jednoduchou úlohu, ve které experimentátor opakovaně schovává hračku pod poklopem A; úkolem dítěte je po každém takovém schování hračky se po ní natáhnout na místo, kam vidělo experimentátora hračku schovat. Po několikerém schování hračky pod poklop A experimentátor schová hračku pod poklopem B. Při minimální časové prodlevě mezi schováním hračky a možností se po ní natáhnout a uchopit ji do rukou se děti staré 8-10 měsíců v této části úkolu často dopouštějí té chyby, že místo toho, aby hledali hračku pod poklopem B, ji hledají pod poklopem A. Děti starší 12 měsíců se takové chyby již nedopouštějí a schovanou hračku správně hledají pod poklopem B. Podle Piageta je příčinou této dramatické vývojové změny to, že děti starší 12 měsíců již disponují plně vyvinutým pojmem stálosti objektu. Jinými slovy, jsou již schopny vytvářet abstraktní mentální reprezentace objektů vnějšího světa. Mladší děti něčeho takového podle Piageta ještě schopny

nejdou – součástí objektu, který reprezentují ve své mysli, je vždy to, co právě vnímají a co právě dělají; objekty vnějšího světa tak pro ně ještě nemají svou vlastní, objektivní a na vnímání a činnosti subjektu nezávislou existenci.

Na Piagetově experimentálním úkolu je zvláštní a zajímavé to, že pozorované chování dětí - tj. to, zda vykazují chybu A-ne-B, či nikoli - je extrémně citlivé na změny v experimentálním designu, tzn. že i velice drobnými změnami a úpravami v uspořádání Piagetova základního experimentálního úkolu lze docílit toho, že děti mladší 12 měsíců budou vůči chybě A-ne-B imunní a naopak že děti starší 12 měsíců se budou chyby A-ne-B dopouštět, přestože by již měli – alespoň podle Piageta – disponovat plně vyvinutým konceptem stálosti objektu. Ve světle takto vysoké míry variability v chování dětí v závislosti na detailech experimentálních podmínek se zdá být Piagetovo vysvětlení dosti problematické a diskutabilní. Objevila se proto celá řada nových hypotéz snažících se vysvětlit chování dětí za různých alternativních experimentálních podmínek.

Tyto nové hypotézy většinou vycházejí z informačně procesního paradigmatu, v rámci kterého je člověk pojímán jako výpočetní systém, který prostřednictvím smyslových orgánů přijímá informace z vnějšího prostředí, vytváří vnitřní symbolické (mentální) reprezentace těchto informací a následně na ně aplikuje různé výpočetní (mentální) operace, prostřednictvím kterých s těmito vnitřními reprezentacemi různě manipuluje a transformuje je na výstupní informace, které mohou mít podobu nějakého úsudku, rozhodnutí, plánu nebo nějaké motorické činnosti. Vývoj schopností dítěte – pojímaného jako druh výpočetního systému - je pak dán nárůstem kapacity různých výpočetních zdrojů, kterými disponuje: paměti, pozornosti, mentálních reprezentací, mentálních procesů atp. Jedna z informačně procesních hypotéz vysvětlujících chybu A-ne-B vychází z předpokladu, že děti, které se v Piagetově experimentálním úkolu dopouštějí chyby A-ne-B, nemají problém s konceptem stálosti objektu, ale se způsobem, jakým reprezentují ve své mysli prostor: Příčinou kognitivního omylu A-ne-B je podle této hypotézy to, že děti mezi 8. a 10. měsícem věku reprezentují okolní prostor „egocentricky“, tj. vždy relativně ke své pozici v prostoru, nikoli objektivně, jakoby z ptačí perspektivy, kde má vše své pevné místo. Takové dítě když bude mít opakovanou zkušenost, že nějaký objekt se vždy nachází nalevo od něj, bude mít potom tendenci tento objekt vždy hledat nalevo od sebe a to bez ohledu na změnu své aktuální pozice v prostoru. Takže jestliže dítě bude několikrát za sebou natahovat svou ruku po hračce směrem doleva, potom bude i při dalších pokusech s vyšší pravděpodobností natahovat svou ruku směrem k místu, které je vůči němu nalevo. Tuto hypotézu se zdá potvrzovat experimentální pozorování, že když dítě sleduje experimentátora schovávajícího hračku pod

poklop B na jedné straně stolu, avšak po hračce má možnost se natáhnout z opačné strany stolu, potom dítě neomylně šahá směrem k poklopu B, neboť ten se nyní nachází vůči dítěti ve stejné relativní poloze jako poklop A na druhé straně stolu. Takže to, co na jedné straně stolu vedlo k chybnému úkonu, na druhé straně stolu vedlo ke správnému rozhodnutí. Podle další informačně procesní hypotézy je příčinou selhání dětí v Piagetově experimentálním úkolu nedostatečně efektivní kódování informací do jejich paměti, takže ty v ní potom nedokážou dostatečně dlouho podržet vzpomínku na místo, kam experimentátor hračku schoval. Tato hypotéza se zdá být v souladu s experimentálním pozorováním, že když se dítěti mezi 8. a 10. měsícem věku umožní, aby se natáhlo po hračce v několika málo vteřinách poté, co ji experimentátor schová pod poklop B, natáhne se správným směrem, tj. k poklopu B; tendenci sahat po poklopu A mají takto staré děti pouze tehdy, když je prodleva mezi schováním hračky a aktivitou dítěte dostatečně dlouhá na to, aby se jim informace o místě uložení hračky mohla z paměti vytratit. Další informačně procesní hypotézy o příčinách chyby A-ne-B svá vysvětlení zakládají na ne/schopnosti inhibice vlastních behaviorálních reakcí, která je spojená s ne/zralostí prefrontální kůry, na míře stability mentálních reprezentací atd. (Smith, Thelen, 2003; Beer, 2000).

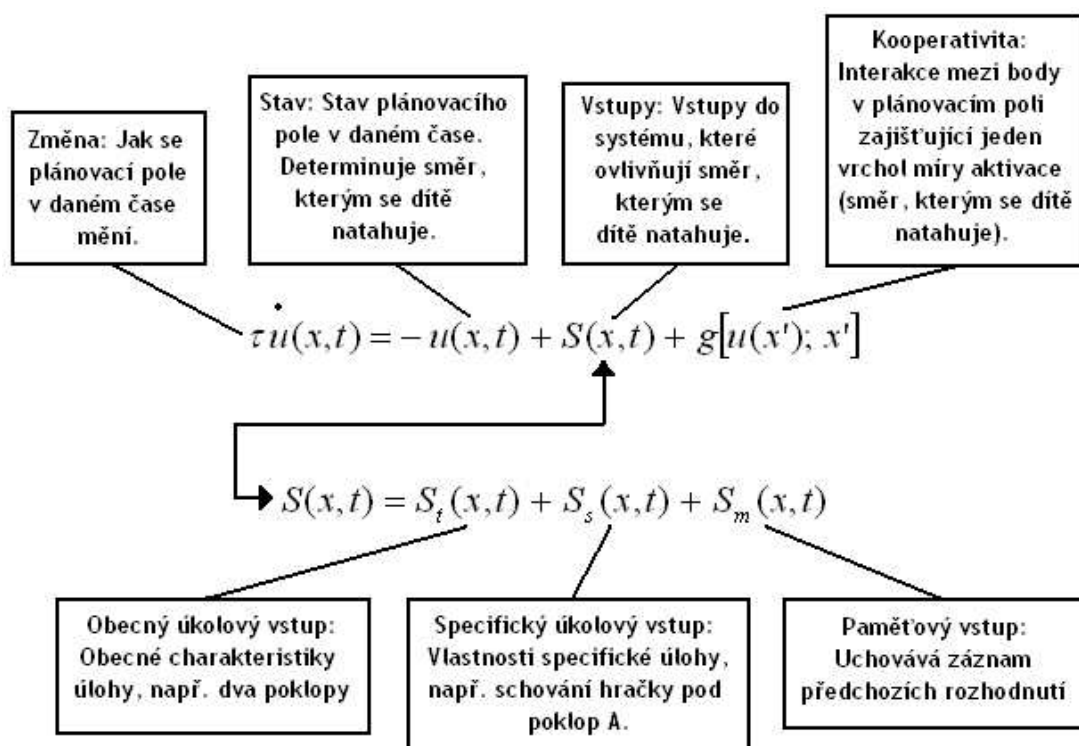
Všem výše uvedeným hypotézám (včetně té Piagetovy) o příčinách chyby A-ne-B je společné to, že postulují existenci jediné příčiny a že tuto příčinu lokalizují dovnitř výpočetního mechanismu lidské mysli. Právě tato monokauzalita je důvodem, proč tyto hypotézy dokáží vysvětlit jen velice omezenou část nashromážděných experimentálních dat. Zcela odlišný přístup proto zvolili Thelen a kol. (2001), kteří vytvořili dynamický **model plánovacího pole**, ve kterém je chyba A-ne-B emergentním důsledkem vzájemného působení a interakce několika různých příčinných faktorů, které se nenacházejí pouze uvnitř mysli dítěte, ale zahrnují také celkovou senzomotorickou interakci dítěte s vnějším prostředím.

Při řešení Piagetova experimentálního úkolu stojí dítě před rozhodnutím, jakým směrem má natáhnout svou ruku, zda doprava, nebo doleva. Tendence dítěte šahat směrem doprava, či doleva se přitom v čase průběžně mění. Soubor těchto tendencí Thelen a kol. nazývají *plánovací pole* (*planning field*), které se velice volně inspiroje funkcí nervového systému: Plánovací pole definuje pomyslný jednodimenzionální prostor možných aktivačních stavů, ve kterých se může nacházet systém, který je odpovědný za plánování a realizaci uchopovací aktivity dítěte. Chování tohoto aktivačního (plánovacího) pole je dáno funkcí $u(x,t)$, která specifikuje, v jakém směru (x) má dítě v daném čase (t) tendenci se natahovat; u se přitom mění v závislosti na řadě různých faktorů (parametrů či vstupů) jako jsou:

- aktuální (aktivační) stav plánovacího pole,

- obecná experimentální situace, tj. přítomnost dvou vizuálně nápadných poklopů na místech A a B,
- specifická experimentální situace, tj. vizuální zkušenost toho, jak experimentátor schovává hračku buď pod poklop A, nebo pod poklop B,
- paměť předchozích rozhodnutí či epizod uchopovací aktivity,
- kompetitivní (inhibiční) interakce mezi jednotlivými body jednodimenzionálního aktivačního prostoru plánovacího pole (umožňující dítěti dojít k jednoznačnému rozhodnutí, kam šáhnout)...

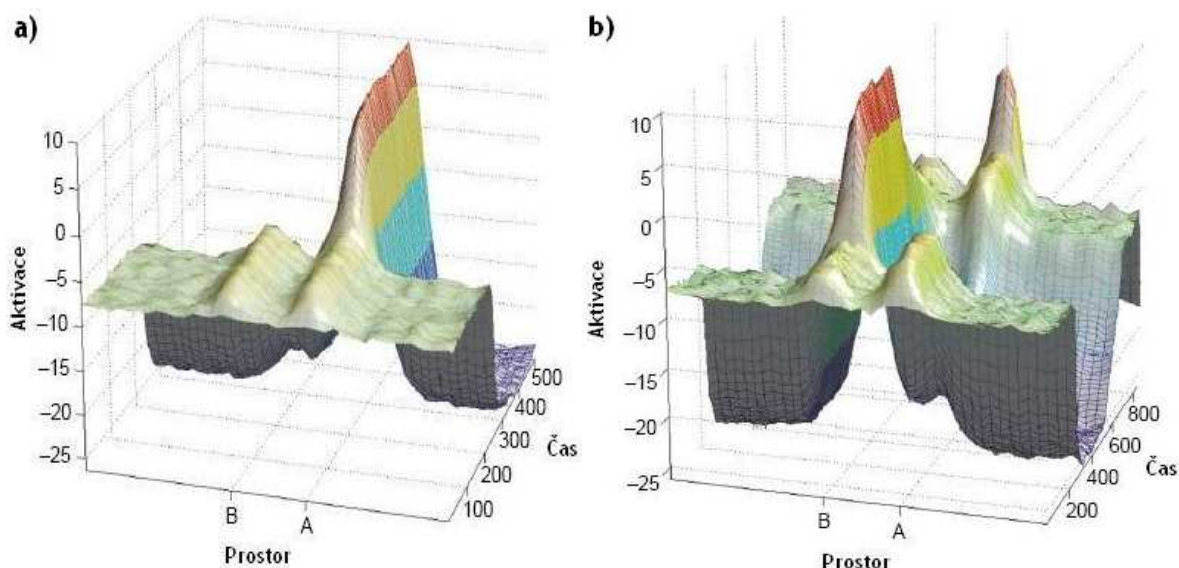
Tyto vstupy jsou vždy charakterizovány tím, na jaké místo plánovacího pole (A/B) působí a jakým způsobem na něj působí (zvýšením/snížením míry aktivace na daném místě plánovacího pole). Chování každého faktoru v čase je specifikováno jeho vlastní funkcí; jejich propojením pak vzniká složitá diferenciální rovnice, která specifikuje jak se u v daném časovém okamžiku mění (viz obrázek 107).



Obrázek 107: Diferenciální rovnice specifikující chování dynamického modelu plánovacího pole. (převzato z Thelen a kol., 2001)

Na obrázku 108 je zachycen vývoj míry aktivace v plánovacím poli během Piagetova experimentálního úkolu: Před tím, než experimentátor hračku schová pod poklop A, dítě ve

svém zorném poli vidí dva vizuálně nápadné poklopy; tomuto vstupu odpovídá mírně zvýšená aktivace v plánovacím poli jak na místě A tak i na místě B (viz obrázek 108a). Poté co experimentátor hračku schová pod poklop A, způsobí tím přechodně zvýšenou míru aktivace v plánovacím poli dítěte na místě A; díky inhibičním interakcím mezi jednotlivými body plánovacího pole také dojde k deaktivaci plánovacího pole na místě B (viz obrázek 108a). Když míra aktivace dosáhne určité kritické (prahové) hodnoty, dítě se na natáhne po hračce schované pod poklopem A. Vzpomínka na tuto uchopovací aktivitu se pak stává vstupem do plánovacího pole při dalším pokusu najít schovanou hračku. Díky této vzpomínce je míra aktivace v plánovacím poli o něco vyšší na místě A než na místě B. Tato zvýšená míra aktivace tvoří paměť všech předchozích epizod uchopovací aktivity, která byla nasměrovaná k poklopu A. Na obrázku 108b je pak zachycena situace, kdy experimentátor po několikerém schování hračky pod poklop A schová hračku pod poklop B: Díky vizuálnímu vstupu dojde



Obrázek 108: Vývoj míry aktivace v plánovacím poli během Piagetovy experimentální úlohy, kdy má dítě za úkol hledat hračku schovanou pod jedním ze dvou poklopů (označených zde písmeny A a B); podrobnější výklad viz hlavní text. (s mírnými úpravami převzato ze Smith, Thelen, 2003)

nejdříve ke zvýšení míry aktivace v plánovacím poli na místě B; po nějaké době však v plánovacím poli v důsledku paměťového záznamu předchozích uchopovacích aktivit zcela převládne aktivita na místě A. Model tak správně předpovídá, že bezprostředně po schování hračky jsou děti schopné hledat hračku na správném místě a že chyby A-ne-B se děti dopouští pouze tehdy, když časová prodleva mezi schováním hračky a uchopovací aktivitou překročí určitou kritickou hranici. Specifickým nastavením jednotlivých parametrů (odpovídajících

různým experimentálním designům) diferenciální rovnice popisující chování plánovacího pole je možné reprodukovat všechna experimentální pozorování spojená se zkoumáním chyby A-ne-B; vedle úspěšné reprodukce všech existujících experimentálních dat tento model navíc dokázal vygenerovat řadu přesných předpovědí týkajících se chování dětí za různých, ještě nerealizovaných experimentálních podmínek. Ukázalo se, že s pomocí tohoto modelu má experimentátor téměř neomezenou možnost libovolně „zapínat“ a „vypínat“ chybu A-ne-B prostřednictvím úprav různých parametrů experimentální situace – změnou délky časové prodlevy mezi schováním hračky a možností se po ní natáhnout, změnou vizuální přitažlivosti poklopů nebo samotného schovávání hračky pod poklop, zvýšením či naopak snížením počtu uložení hračky pod poklop A atd.

Na rozdíl od tradičních modelů kognitivních procesů se dynamický model plánovacího pole při vysvětlování chyby A-ne-B nezaměřuje pouze na to, co se odehrává uvnitř dětské mysli, na jeho mentální reprezentace, na mentální procesy, kterými tyto reprezentace zpracovává, ale na celkovou činnost dítěte v kontextu specifických podmínek vnějšího prostředí: Podle tohoto modelu je chyba A-ne-B emergentním důsledkem průběžné a v čase se vyvíjející interakce mezi souborem různých proměnných, které se nacházejí jak na straně dítěte, tak na straně vnějšího prostředí, takže to, jestli dítě daný úkol zvládne, či nikoli nezávisí pouze na jeho vnitřních (výpočetních) zdrojích, ale z velké části také na řadě parametrů vnějšího prostředí, se kterým je dítě v interakci a ve kterém je jeho aktivita zakotvena (*embedded*).

Na význam vnějšího prostředí pro pochopení lidské kognice upozorňoval také Herbert Simon (1981), který v této souvislosti použil přirovnání snahy porozumět kognitivním procesům člověka ke snaze porozumět kognitivním procesům mravence pohybujícího se po pláži: Pohyb takového mravence po pláži může být i poměrně složitý a vnějšímu pozorovateli by se tak mohlo zdát, že aby byl mravenec schopen tak složitého chování, musí být vybaven nějakým sofistikovaným (biologickým) výpočetním systémem, který by si dokázal poradit s vytvářením mentálního modelu okolního prostředí a s plánováním a řízením jeho pohybů založeným na tomto vnitřním modelu. Problematicnost tohoto závěru je podle Simona v tom, že složitý pohyb mravence nám toho mnohem více říká o složitosti tvaru či struktury pláže než o složitosti nějakého výpočetního zařízení řídícího pohyby mravence, který ve skutečnosti pouze jednoduše reaguje na tvar a strukturu pláže.¹ Poučení, které z toho podle Simona

¹ “Viewed as a geometric figure, the ant’s path is irregular, complex, hard to describe. But its complexity is really a complexity in the surface of the beach, not a complexity in the ant.” (Simon, 1981, s. 51)

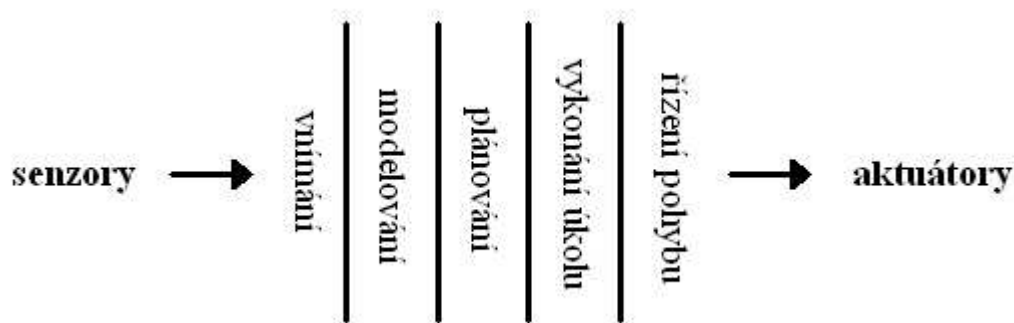
vyplývá, je že **by se neměla podceňovat role prostředí v ovlivňování a tvarování chování člověka nebo jiných kognitivních systémů.**

Ze sociální psychologie je dobře známo (Výrost, Slaměník, 1997), že lidé mají obecně tendenci podceňovat vliv vnějších situačních podmínek na chování a předpokládat, že chování je především odrazem určitých vnitřních dispozičních charakteristik jedince (tzv. *základní atribuční chyba*). Obdobnou tendenci bylo až do relativně nedávné doby možné nalézt také v přístupu kognitivních vědců k předmětu svého zkoumání. Kognitivní vědci vnější prostředí většinou vnímali jako něco, co je zdrojem problému specifikujících podnětů a co je arénou pro realizaci vnitřně vygenerovano plánu jejich řešení. Takto pojímané prostředí hrálo v teoriích kognitivních procesů roli pouze pasivního zdroje informací a pasivního objektu působení aktivního kognitivního systému, především na jehož (vnitřní, výpočetní/kognitivní) činnost se upírala pozornost kognitivních vědců. Teprve někdy od poloviny 80. let minulého století se začalo větší množství kognitivních vědců dívat na prostředí jako na něco, co by se mohlo samo aktivně spolupodílet na generování adaptivního chování kognitivních systémů. Tato změna pohledu na roli vnějšího prostředí měla do značné míry svůj zdroj v **robotice** – v oboru umělé inteligence, který se snaží vytvářet mobilní autonomní roboty (*moboty*) schopné se samostatně pohybovat a plnit některé jednoduché úkoly v reálném prostředí skutečného světa, ve kterém roboty musí být schopny rychle, flexibilně a v reálném čase reagovat na neustále se měnící podmínky a nepředvídatelné situace.

Jak již zde bylo několikrát uvedeno, to, co podle tradiční kognitivní vědy tvoří podstatu kognice, jsou výpočetní procesy manipulace s vnitřními symbolickými reprezentacemi vnějšího světa. Ze stejného předpokladu vycházela i tradiční robotika, podle níž robotický systém prostřednictvím svých senzorů přijímá z vnějšího prostředí vstupní informace, ze kterých si ve své mysli vytváří vnitřní model či symbolickou reprezentaci světa; na tomto vnitřním symbolickém modelu následně provádí určité výpočetní operace, které vedou k výstupu v podobě nějaké specifické behaviorální reakce. Jozef Kelemen (1994, s. 47) chování takového robota popisuje jako „*výsledek zvažování možností, plánování akcí na základě vnitřní představy (reprezentace) stavů světa, vlastních schopností... působit na svět a na základě anticipace účinků svých budoucích úkonů.*“ Funkční architektura tohoto tzv. **kognitivního robotického systému** se skládá ze tří základních podsystémů:

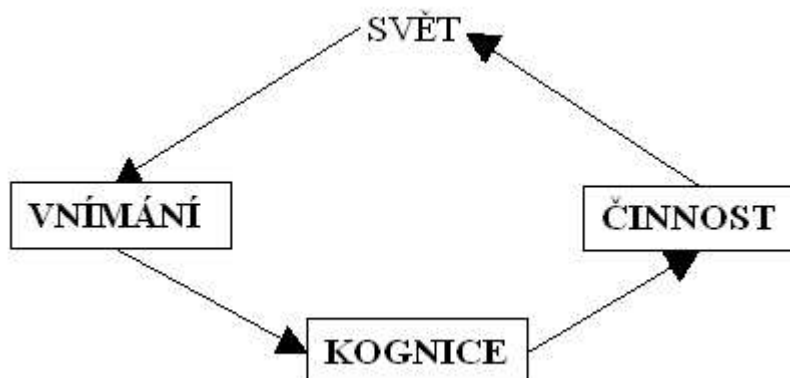
- *vnímacího podsystému*, který robota prostřednictvím jeho smyslů (senzorů) informuje o aktuálním stavu okolního prostředí,
- *motorického podsystému*, který robotovi prostřednictvím jeho aktuátorů umožňuje se pohybovat ve světě nebo na něj nějak jinak fyzicky působit, a

- *kognitivního pod systému*, který robotovi jednak umožňuje si vytvářet vnitřní reprezentace vnějšího prostředí, svých možností v něm jednat a souvislostí mezi svou činností a jejími důsledky a jednak na základě těchto vnitřních reprezentací plánovat svou činnost a aktivitu ve vnějším prostředí.



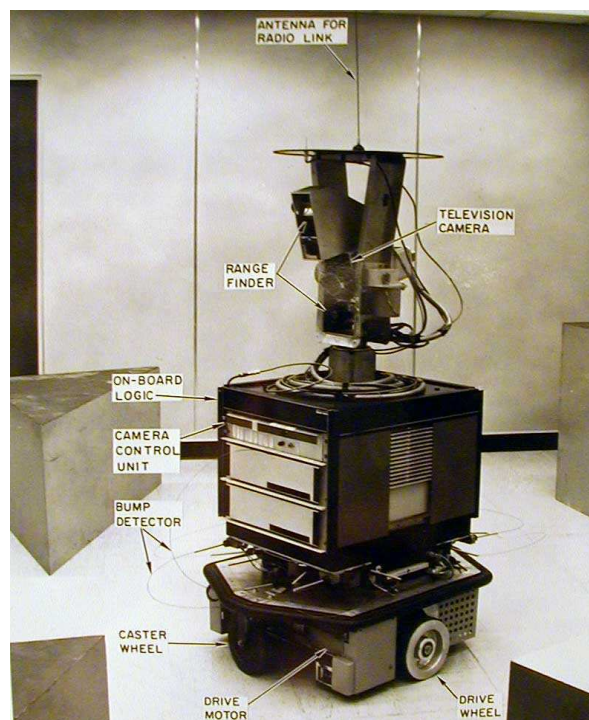
Obrázek 109: Tradiční dekompozice řídicího systému mobilního robota na jednotlivé funkční moduly. (s mírnými úpravami převzato z Brooks, 1999, s. 4)

Robotik *Rodney Brooks* v této souvislosti hovoří o modulu kognice (*cognition box*), který zprostředkovává interakci mezi modulem vnímání (*perception box*) a modulem činnosti (*action box*). Slabinou této funkční architektury je to, že vytváření a manipulace detailní



Obrázek 110: Tradiční model kognitivního robotického systému, ve kterém kognice tvoří spojovací článek mezi vnímáním vnějšího světa a plány činností ve vnějším světě (s mírnými úpravami převzato z Brooks, 1999, s. viii)

vnitřní reprezentace vnějšího prostředí je výpočetně velice náročný proces, který klade extrémně vysoké nároky na kapacitu výpočetních zdrojů, kterými robotický systém disponuje. V praxi to znamená, že takové robotické systémy nejsou schopny v reálném čase reagovat na proměny okolního prostředí. Klasickým příkladem slabin tohoto tradičního přístupu je *Shakey* – jeden z prvních autonomních mobilních robotů na světě postavený v letech 1966-1972



Obrázek 111: Shakey. Jeden z prvních autonomních mobilních robotů postavený v letech 1966-1972 vědci ze Stanfordského výzkumného institutu v Kalifornii.

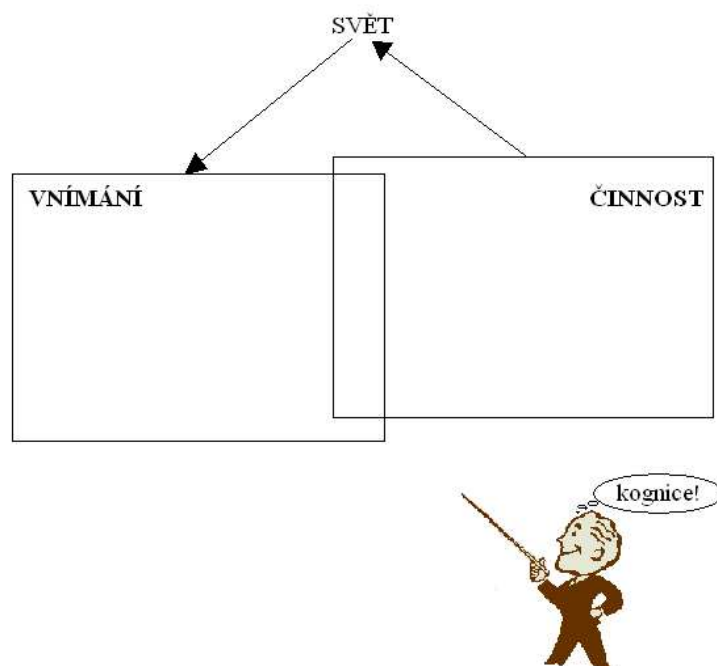
týmem vědců ze Stanfordského výzkumného institutu v Kalifornii. Shakey (viz obrázek 111) byla vysoká krabice na kolečkách, ke které byla připojena kamera a vzdálený (na tehdejší dobu vysoce výkonný) počítač. Shakey byl svými tvůrci dopředu vybaven základním modelem prostředí, ve kterém se měl pohybovat, a sofistikovaným softwarem, který mu umožňoval tento vnitřní model průběžně aktualizovat a plánovat na jeho základě své pohyby.¹ Pro Shakeyho bylo charakteristické, že na vnímání svého okolí a na plánování svých pohybů potřeboval velké množství času: Shakey nejdříve otevřel své pomyslné oči, podíval se na svět před sebou a ve své „mysli“ si vytvořil jeho vnitřní model, což mohlo trvat i několik minut. Poté Shakey konzultoval tento svůj vnitřní model okolního prostředí s cílem vygenerovat a následně vyhodnotit přijatelnost různých alternativních akcí. Po naplánování své cesty se nakonec posunul o necelý metr vypočítaným směrem. Potom se celý cyklus opakoval: otevřít oči, podívat se na svět, vytvořit si nový vnitřní model světa, naplánovat svůj pohyb a posléze ho zrealizovat. Každý pohyb, který kdy Shakey udělal, byl přitom svým způsobem krokem do tmy, neboť mezi jeho aktem vnímání a fyzickou realizací pohybu vždy uběhlo několik minut. Shakeyho úspěšné navigování ve světě tak bylo kriticky závislé na předpokladu, že svět se

¹ Shakey k reprezentaci svého světa (sestavajícího z větší místnosti s rovnou podlahou, na které bylo několik krabic, kterým se Shakey musel vyhýbat) používal predikátovou logiku prvního řádu, která mu umožňovala reprezentovat svět dostatečně detailně na to, aby s pomocí své vnitřní reprezentace mohl bezpečně plánovat své pohyby v tomto světě (Brooks, 2003, s. 38).

v průběhu jeho rozvažování o tom, co dál podniknout, nebude nijak významně měnit. Na první pohled je zřejmé, že tímto způsobem fungující robot by v podmínkách reálného světa, kde se situace mění každým okamžikem, nemohl obstát. Příčinou toporného pohybu Shakeyho je výpočetně (a časově) vysoce náročný proces vytváření a manipulace detailní vnitřní reprezentace okolního prostředí. Avšak jestliže jsou zdrojem potíží vnitřní reprezentace, tak proč se jich nezbatit? A přesně to navrhl Rodney Brooks (1999), který našel inspiraci v živočišné říši, konkrétně u hmyzu, který je schopen přesné a rychlé navigace ve složitém prostředí, přestože obvykle disponuje maximálně několika stovkami neuronů, které v žádném případě nemohou být schopny realizovat vysoce komplexní výpočetní algoritmy, které robotici jinak obvykle navrhují pro své kognitivní robotické systémy.

Vzhledem ke svým omezeným výpočetním zdrojům musí tedy hmyz využívat jinou a mnohem efektivnější strategii než je budování a manipulace detailní vnitřní reprezentace okolního prostředí. Podle Brookse tato strategie spočívá v tom, že se světu přenechává role jeho vlastního modelu a paměti, neboť **svět je sám sobě tím nejlepším (a „nejlevnějším“)** **modelem**. Robotický systém založený na této strategii tak nemusí vytvářet ve své „mysli“ žádnou další kopii okolního prostředí, aby se v něm dokázal pohybovat, neboť roli vnitřního modelu podílejícího se na generování a formování chování robotického systému převzalo samotné vnější prostředí. Aby svět mohl hrát tuto roli vnější paměti, je nutné, aby robotický systém měl ke světu rychlý a okamžitý přístup kdykoli je to zapotřebí. Této flexibility v přístupu k informacím obsaženým ve vnějším prostředí lze dosáhnout tím, že se robotickému systému umožní jednoduše, rychle a bez nějakého velkého prodlení reagovat na informace přicházející k němu z vnějšího prostředí prostřednictvím jeho senzorů. Takový robotický systém se skládá de facto pouze ze subsystému vnímání a subsystému činnosti, mezi kterými není žádný zprostředkovatel (v podobě subsystému kognice). Model robotického systému založeného na této myšlence je zobrazen na obrázku 112: Jediné, co obsahuje, jsou moduly vnímání a činnosti, které jsou bezprostředně spouštěny signály přicházejícími z vnějšího prostředí; kognice je zde pak něco, co existuje pouze v oku pozorovatele, kterému se zdá, že systém vykazující komplexní inteligentní chování v sobě musí obsahovat i nějaký modul kognice (zajišťující mapování okolního prostředí a plánování akcí), přestože ve skutečnosti je toto chování pouze emergentním důsledkem interakce souboru jednoduchých reakcí spouštěných podněty z vnějšího prostředí: „*Uvědomili jsme si,*

že tzv. centrální systémy inteligence... jsou možná zbytečnou iluzí a že veškerá... inteligence vyrůstá z propojení percepce a aktuátorů.“¹ (Brooks, 1999, s. viii)



Obrázek 112: Model tzv. reaktivního robotického systému, ve kterém skutečně existují pouze moduly vnímání a činnosti; kognice zde existuje pouze v oku vnějšího pozorovatele (s mírnými úpravami převzato z Brooks, 1999, s. xi)

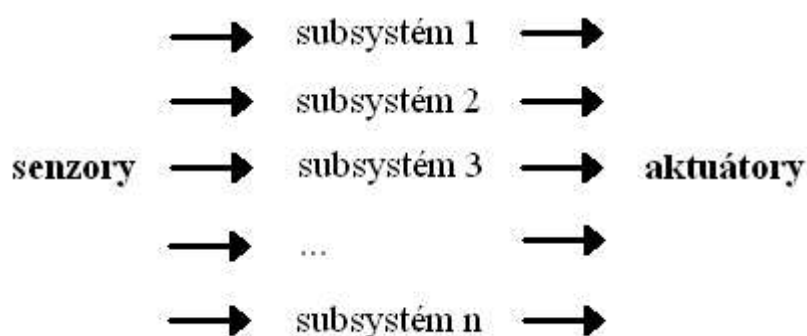
Brooks tuto obecnou myšlenku podrobněji rozpracoval do podoby návrhu tzv. **subsumpční architektury**: Robotický systém mající tento typ funkční architektury se skládá z několika relativně nezávislých jednoduchých subsystémů, z nichž každý zajišťuje určitou specifickou aktivitu. Úkolem těchto subsystémů není vytvářet či předávat nějaký druh symbolických reprezentací vstupů z vnějšího prostředí, ale zajišťovat rychlou přeměnu vstupu na konkrétní akci. Komunikace mezi jednotlivými subsystémy je přitom omezena na několik málo velice jednoduchých signálů, které jednotlivým subsystémům umožňují například vzájemně tlumit nebo naopak podporovat svou aktivitu, nikoliv však již si předávat nějaké komplexnější informace. Jednotlivé subsystémy lze na sebe postupně vrstvit a vytvářet tak čím dál tím sofistikovanější robotické systémy schopné stále komplexnějšího chování. Brooksovi se tímto způsobem podařilo vytvořit plně autonomní robotický systém (nazývaný *Allen*) ze tří vrstev jednoduchých subsystémů (Brooks, 2003, s. 40-41):

¹ “The realization was that the so-called central systems of intelligence... was perhaps an unnecessary illusion, and that all the power of intelligence arose from the coupling of perception and actuation systems.”

Vrstva 1: „Vyhýbání se objektům“ s pomocí prstence ultrazvukových sonarových senzorů. Tento subsystém způsobuje, že se mobot zastaví, kdykoli se přímo před ním objeví nějaká překážka, a že se natočí směrem, kde se žádné překážky nenacházejí.

Vrstva 2: „Bezcílné toulání“. V případě, že není subsystém ve vrstvě 1 aktivní, tento modul generuje náhodné změny směru jízdy, čímž mobotovi umožňuje bezcílně prohledávat své okolí. Díky tomu, že subsystém v nižší vrstvě 1 zajišťuje předcházení kolizím, může být tento subsystém velice jednoduchý, neboť se nemusí nijak starat o riziko případného střetu s nějakou překážkou.

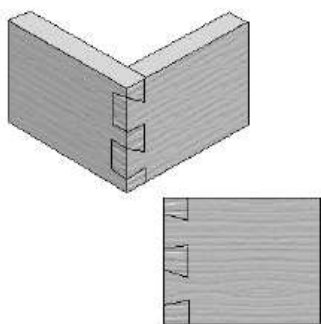
Vrstva 3: „Sledování“. Když tento subsystém objeví v okolí něco zajímavého, utlumí činnost subsystému ve vrstvě 2 a nasměruje pohyb mobota směrem k zajímavému objektu. Ani tento subsystém se díky subsystému ve vrstvě 1 nemusí starat o vyhýbání se překážkám; v případě, že se v okolí žádný zajímavý objekt nevyskytuje, přebírá aktivitu subsystém v nižší vrstvě 2, který zajišťuje náhodné prohledávání okolí, čímž se zvyšuje pravděpodobnost, že mobot narazí na něco zajímavého.



Obrázek 113: Dekompozice reaktivního mobota se subsumpční architekturou na jednotlivé subsystémy. (s mírnými úpravami převzato z Kelemen, 1994, s. 73)

Důležitou vlastností subsumpční architektury je to, že v ní neexistuje žádné privilegované místo, žádná centrální databáze či procesor, kde by se v podobě nějakého symbolického kódu shromažďovaly informace z jednotlivých subsystémů a kde by se tyto informace zpracovávali za účelem koordinace činnosti jednotlivých subsystémů. Robotický systém se subsumpční architekturou je spíše souborem konkurujících si subsystémů, jejichž činnost je (vedle minimální vnitřní komunikace) koordinována především signály přicházejícími z vnějšího prostředí: Neexistuje zde žádné centrální řízení ani celkový plán činnosti; místo toho prostředí samo řídí chování robotického systému prostřednictvím souboru jeho vlastních základních behaviorálních reakcí, které se spouštějí přijetím specifických signálů z vnějšího prostředí.

Celkové chování robotického systému tak vyvstává z neustálé interakce mezi behaviorálními reakcemi robota a jeho prostředím. Schopnosti robotického systému takto tvoří s prostředím těsně propojený systém – jakýsi rybinový spoj (viz obrázek 114) -, ve kterém jsou funkční charakteristiky robota dokonale přizpůsobeny charakteristikám prostředí, ve kterém se pohybuje. Inteligence tu tak není nějakou individuální vlastností mobota, ale je spíše označením určitého specifického druhu interakce mobota s jeho prostředím, nebo ještě lépe řečeno, je charakteristikou vztahu mobota ke svému prostředí. Toto pojetí inteligence má



blízko k Brunswikovu širšímu „pojetí účelu psychiky jako prostředku, který pomáhá organismu při adaptaci na změny prostředí – tedy přežít“; z čehož pak Brunswik vyvozuje, že „psychologie by měla studovat vztahy mezi organismem a prostředím“, resp. že by se měla „zabývat prostředím stejně jako organismem, jinak nemůže porozumět jejich vzájemnému vztahu a souvislostem.“ (Kostroň, 1997, s. 18)

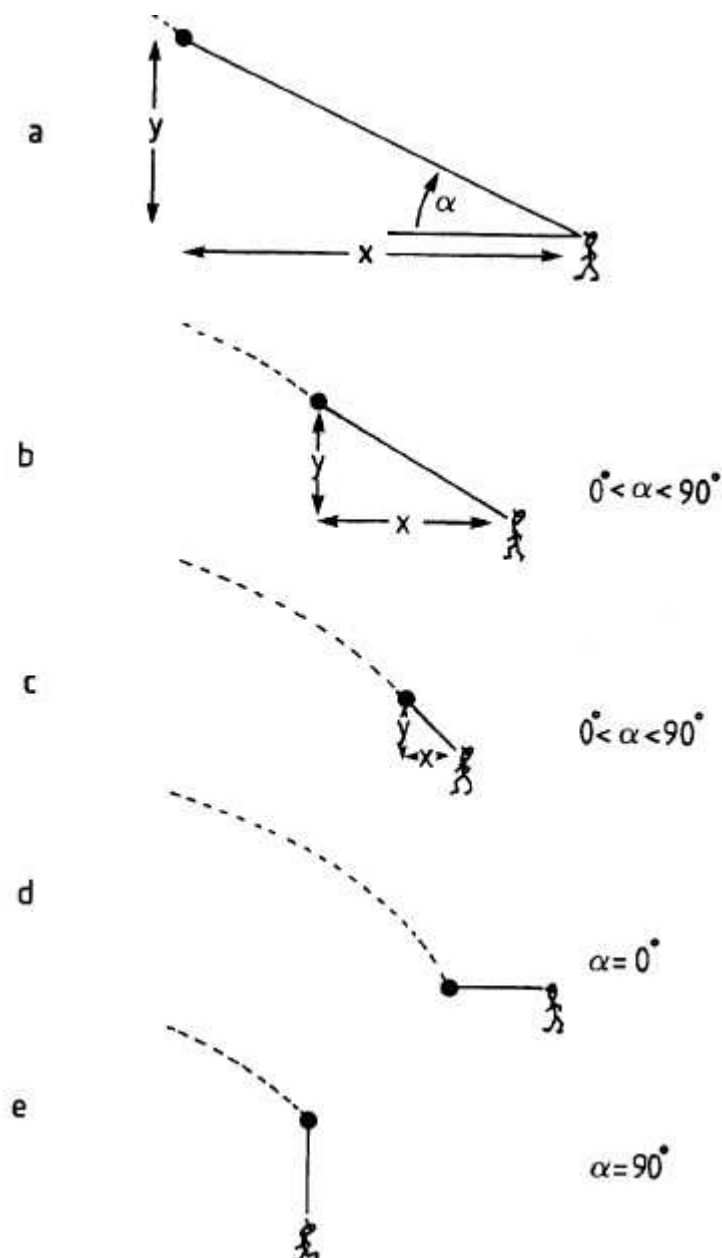
Obrázek 114: Rybinový spoj.

Příkladem dokonalého sladění a spolupráce schopností mobota a charakteristik jeho prostředí je *Herbert* – další mobot z dílny Rodneyho Brookse. Úkolem Herberta bylo sbírat ze stolů v laboratoři prázdné plechovky od limonád, což je poměrně obtížný úkol, který vyžaduje schopnost pohybovat se v prostředí, které se neustále mění, kde je plno překážek, kterým je potřeba se vyhnout, a kde je nepřeberné množství různých věcí, mezi nimiž je třeba identifikovat stoly a na nich ležící plechovky. Ke splnění daného úkolu by tradiční UI mobota vybavila schopností vytvářet si ve své „mysli“ detailní model svého prostředí a schopností na základě tohoto modelu plánovat svou činnost. Problematicnost tohoto přístupu je v tom, že je příliš výpočetně a časově náročný. Herbert byl proto postaven na základě principů subsumpční architektury: Skládal se z několika vrstev jednoduchých subsystémů, z nichž každý odpovídal za nějaké jednoduché, ale kompletní chování. Základní subsystémy zajišťovaly pohyb a vyhýbání se překážkám; tyto subsystémy přerušili svou činnost, jakmile Herbertův primitivní zrakový systém identifikoval objekt, jehož obrys se podobal obrysu stolu. Když se Herbert dostal ke stolu, subsystémy zajišťující pohyb a vyhýbání se překážkám předaly kontrolu subsystémům zajišťujícím prozkoumávání stolu pomocí videokamery a laseru. Jakmile tyto subsystémy identifikovaly základní obrys plechovky, tělo Herberta se okolo své osy natáčelo tak dlouho, dokud se plechovka nenacházela v centru jeho zorného pole. To pak byl signál pro zastavení otáčivého pohybu a pro aktivaci robotické paže, která

byla vybavena jednoduchými tlakovými senzory. Jakmile Herbert ucítil specifický „plechovkovitý“ tvar, subsystém odpovědný za uchopování zajistil sklizení plechovky ze stolu (Brooks, 2003). Navzdory absenci jakéhokoli „centrálního plánovače“ Herbert takto dokázal splnit svůj úkol čistě prostřednictvím jednoduchých behaviorálních reakcí, nad jejichž spouštěním a koordinací nemusel vůbec uvažovat nebo přemýšlet, neboť to za něj dělalo samotné vnější prostředí, na které byl adaptován. Důležité je zde poznamenat, že Herbert nebyl pouze pasivním objektem signálů přicházejících k němu z vnějšího prostředí, neboť Herbert sám svou aktivitou do značné míry ovlivňoval to, jaké signály, kdy a v jakém pořadí se k němu dostanou. Příkladem zde budiž jeho centrování plechovky do středu jeho zorného pole, které bylo signálem pro automatické potlačení nebo naopak spuštění aktivity některého z dalších subsystémů.

Je zajímavé, že podobnou strategii jako Herbert používají také baseballisté při chytání vysokého míčku: Chytání míčku je podobně jako sbírání plechovek ze stolu poměrně náročný úkol. Kdybychom chtěli použít prostředky tradiční UI k sestavení robota-baseballisty, který by něco takového dokázal, museli bychom robota vybavit mnoha informacemi (například o místě, kde se robot právě nachází, o místě, odkud byl míček odpálen, o jeho počáteční rychlosti, směru rotace, rychlosti a směru větru atp.), sofistikovaným softwarem, který by dokázal všechny tyto informace náležitě zpracovat, a nakonec dostatečně výkonným hardwarem, který by umožnil všechny výpočty provést dostatečně rychle na to, aby robot ještě stačil doběhnout na vypočítané místo dopadu a mohl si na míček počkat. Vzhledem k vysokému tempu hry a skutečnosti, že míček se ve vzduchu zdrží v průměru jen asi 2 vteřiny, je jen málo pravděpodobné, že by se takový robot stal platným členem týmu. Jako mnohem schůdnější se jeví být možnost vybavit robota jednoduchou a výpočetně nenáročnou heuristikou „pohledu“ (*gaze heuristic*), kterou podle všeho při chytání vysokých míčků používají skuteční hráči baseballu: *Peter McLeod* and *Zoltan Dienes* (1996) zjistili, že poté, co pálkař odpálí nadhozený míček, hráč v poli zhruba půl vteřiny jen stojí a odhaduje, zda dráha míčku směřuje před něj nebo za něj; potom začne tímto směrem utíkat a přitom neustále udržuje svůj pohled fixovaný na míček letící ve vzduchu. Jednoduchá heuristika pohledu pak spočívá v upravování rychlosti běhu tak, aby úhel mezi okem a míčkem zůstal v určitém konkrétním rozmezí (viz obrázek 115). Dodržování tohoto jednoduchého pravidla hráči zajistí, že se s míčkem před jeho dopadem určitě někde srazí, takže hráči pak stačí nastavit svou rukavici a míček do ní „ulovit“. Výhoda této heuristiky spočívá v tom, že si hráč - potažmo robot - nemusí zjišťovat hodnoty velkého počtu proměnných a hlavně že nemusí provádět mnoho složitých výpočtů: Hráč sleduje jen jednu proměnnou (úhel mezi okem a

míčkem) a díky jednoduché heuristice pohledu se ocitne právě tam, kam míček dopadne, a přitom se vůbec nemusí starat o to, kde přesně to bude. Využitím zákonitostí vnějšího prostředí lze takto přeměnit časově a výpočetně extrémně náročnou úlohu na problém, který lze zvládnout i s omezenými výpočetními zdroji, kterými obvykle disponuje mozek živých organismů, včetně člověka.



Obrázek 115: Heuristika pohledu. α je úhel elevace pohledu hráče sledujícího odpálený míček (a). Jestliže hráč běží rychlostí, která tento úhel udržuje v rozmezí $0^\circ < \alpha < 90^\circ$, potom má hráč jistotu, že se s míčkem srazí ještě před jeho dopadem (b, c). V případě, že se úhel ocitne mimo toto rozmezí, potom míček dopadne buď před hráče (d) nebo za něj (e). (převzato z McLeod, Dienes, 1996)

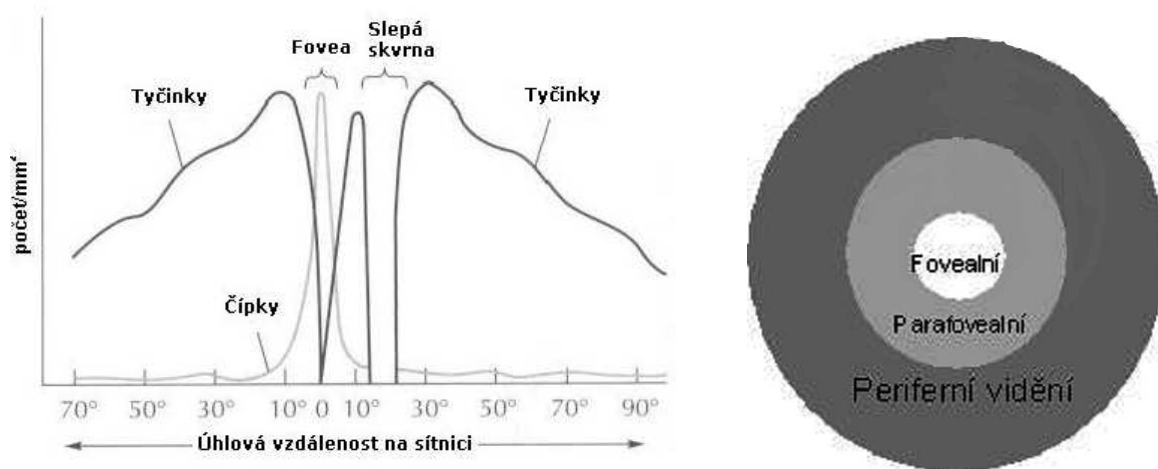
Kognitivní vědec *Gerd Gigerenzer* a kol. (1999, 2000) v této souvislosti hovoří o tzv. **ekologické racionalitě**, jejímž jádrem jsou **rychlé a efektivní heuristiky** (*fast and frugal heuristics*) rozhodování a usuzování, které vyžadují jen minimum času, informací a výpočetních zdrojů, neboť maximálně využívají struktury informací obsažených ve vnějším prostředí. Takto se lze v reálném světě například velice často setkat s tím, že schopnost člověka (roz)poznat nějaký objekt se systematicky (tj. nenáhodně) mění v závislosti na některé jeho důležité vlastnosti (jako je jeho velikost, rychlost, nebezpečnost, kvalita apod.), kterou člověk využívá jako kritérium při svém rozhodování („Jaká značka aut je bezpečnější, Volvo, nebo Geely?“). Této skutečnosti využívá jednoduchá **heuristika rozpoznávání** (*recognition heuristic*), která člověku umožňuje se snadno, rychle a většinou také správně rozhodovat mezi dvěma různými možnostmi z hlediska určitého kritéria - jehož konkrétní hodnotu ovšem člověk nezná - čistě na základě své neznalosti, resp. na základě toho, zda jsou mu posuzované možnosti povědomé, či nikoli. Heuristika rozpoznávání má následující podobu: *„Jestliže jeden ze dvou objektů je rozpoznán a druhý nikoli, potom z toho vyvod', že rozpoznáný objekt má v daném kritériu vyšší hodnotu [nebo nižší, to v případě, že mezi rozpoznáním a daným kritériem existuje negativní korelace]...Heuristiku rozpoznání lze přitom použít pouze tehdy, když člověk není schopen rozpoznat [jen] jeden ze dvou objektů, tedy za situace [pouze] částečné neznalosti.“*¹ (Gigerenzer, Todd a kol., 1999, s. 41) Takto například většinou platí, že čím má nějaké město větší počet obyvatel, tím je vyšší pravděpodobnost, že o něm člověk uslyší a že také bude znát jeho jméno, a naopak, že čím má město menší počet obyvatel, tím bude tato pravděpodobnost nižší; díky existenci této souvislosti a s pomocí heuristiky rozpoznání by měla být velká většina českých studentů (kteří disponují pouze částečnou ne/znalostí amerických reálií²) schopna správně rozhodnout, že ze dvou srovnatelně velkých amerických měst, San Diego a San Antonia, má větší počet obyvatel první jmenované a známější San Diego. Borges, Goldstein, Ortmann a Gigerenzer (1999) demonstrovali efektivitu této jednoduché heuristiky na jejím použití při sestavení fiktivního investičního portfolia sestávajícího z reálných investičních titulů nabízených na

¹ „If one of two objects is recognized and the other is not, then infer that the recognized object has the higher value. [...] The recognition heuristic can only be applied when one of the two objects is not recognized, that is, under partial ignorance.“

² Klíčová je zde jen částečná ne/znalost amerických reálií, neboť při příliš mnoha znalostech a při žádných znalostech nelze heuristiku rozpoznání použít (člověk snadno rozpozná obě nabízené možnosti, resp. nerozpozná ani jednu). Gigerenzer takto ve svém výzkumu zjistil, že zatímco ze zkoumaného vzorku amerických vysokoškolských studentů (u kterých lze předpokládat nadprůměrné znalosti amerických reálií) odpovědělo na otázku „Které město má větší počet obyvatel, San Diego, nebo San Antonio?“ správně pouze 62 % studentů, ze srovnatelného vzorku německých vysokoškoláků (u nichž lze předpokládat jen částečnou obeznámenost s americkými reáliemi) jich na tuto otázku správně odpovědělo plných 100 % (Gigerenzer, Todd a kol., 1999, s. 43).

newyorské a několika německých burzách cenných papírů. Srovnáním výnosu takto sestaveného portfolia (kdy si lidé vybírali ze seznamu 798 investiční titulů na základě známosti jmen jejich emitentů) s výnosy investičních portfolií sestavených s pomocí tradičních nástrojů statistické analýzy se ukázalo, že heuristika dostupnosti představuje plnohodnotný nástroj rozhodování se na finančních trzích, který je z hlediska výnosnosti plně srovnatelný s tradičními a výpočetně mnohem náročnějšími statistickými nástroji.

Míru, v jaké se lidský mozek spoléhá na externí paměť okolního prostředí, dokládá také **nekumulativní povaha sakadických očních pohybů**, která se projevuje **slepotou ke změnám**: Jak známo, na sítnici lidského oka se nacházejí dva základní typy světločivných buněk – tyčinky, které rozlišují odstíny šedi a detekují pohyb, a čípky, které zprostředkovávají ostré a barevné vidění. Tyto buňky nejsou na sítnici rozmístěny rovnoměrně (viz obrázek 116): Zatímco tyčinky lze nalézt prakticky po celé ploše sítnice, čípky se nacházejí téměř výhradně poblíž optické osy, v místě nazývaném **žlutá skvrna (fovea)**, což je velmi malá ploška sítnice (mající tvar prohlubně, odtud také název fovea - jáma), kde se nachází oblast nejostřejšího vidění. Důsledkem tohoto rozložení buněk je, že lidské oko vidí ostře pouze



Obrázek 116: Rozložení tyčinek a čípků na sítnici lidského oka a výsledné typy vidění. (první obrázek převzat z Rains, 2001, druhý obrázek převzat z Klimeš, 2001)

v úhlu 2° zorného pole, což je přibližně velikost dvacetikoruny na vzdálenost natažené paže. Zbylá část sítnice zajišťuje parafoveální a periferní vidění, které je mnohem méně ostré (jeho ostrost je o 15 – 50 % nižší než u vidění foveálního) a také hůře rozlišuje barvy, na druhou stranu však lépe vidí za zhoršených světelných podmínek a také lépe detekuje pohyb. Hlavní



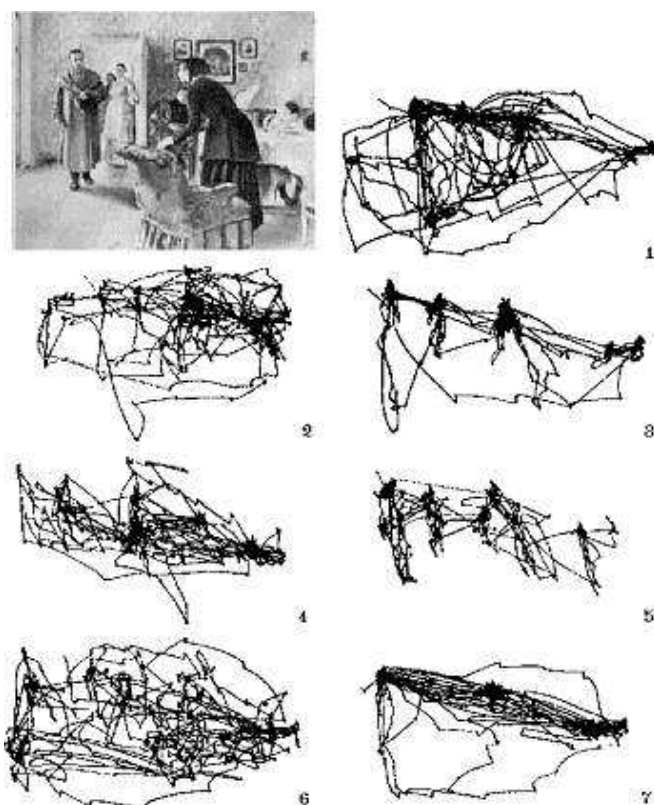
Obrázek 117: *Simulace foveálního, parafoveálního a periferního vidění. (převzato z Miall, Tchalenko, 1998)*

funkcí periferního vidění je předávat mozku informace o zajímavých, důležitých nebo pohybujících se objektech, na které mozek může následně přesunout a fixovat oblast ostrého vidění, aby mu tato zprostředkovala detailnější informace. Tyto tzv. **sakadické oční pohyby** mají dvě základní fáze: fázi fixace, kdy se oko nachází v relativním klidu a „načítá“ informace, a fázi sakád - rychlých pohybů z jednoho místa fixace na další místo fixace. K „načítání“ informací dochází pouze v době fixací, v průběhu sakády je oko prakticky slepé. Doba fixace variuje v závislosti na komplexnosti vnímaného materiálu a na typu úlohy, kterou člověk plní, v průměru však fixace trvají 250-300 ms (Lukavský, 2005; Klimeš, 2001). Yarbus (1967) zjistil, že od typu plněné úlohy odvisí také celkový vzorec sakadických očních pohybů – tzn. že mozek přesouvá oblast ostrého vidění vždy tak, aby získal přístup k těm informacím, které jsou důležité pro vyřešení daného úkolu (viz obrázek 119). Tato zjištění o fungování zrakového systému se zdají naznačovat, že lidský mozek používá sakadické oční



Obrázek 118: *Grafický záznam fixací a sakád. (převzato z Klimeš, 2001)* Na přiloženém CD je v adresáři Záznamy_z_oční_kamery také několik krátkých videozáznamů sakadických očních pohybů získaných z webových stránek firmy DIMAR (<http://www.ocnikamera.cz/>), která se zabývá využíváním údajů z oční kamery k analýze působení reklamního materiálu na chování zákazníků.

pohyby k získávání informací, ze kterých potom vytváří detailní vnitřní reprezentaci okolního prostředí. Tomu se zdá také odpovídat subjektivní pocit, že člověk disponuje bohatou a do všech detailů propracovanou vizuální zkušeností svého okolí. Nedávné výzkumy však ukazují, že tento subjektivní pocit je do značné míry iluzorní a že mozek nevytváří detailní



Obrázek 119: Yarbus prezentoval testovaným osobám jeden obrázek a zkoumal, jak se mění vzorec sákladických očních pohybů v závislosti na úkolu, který tyto osoby dostaly: 1) Volné pozorování, 2) odhad movitosti osob na obrázku, 3) odhad stáří osob, 4) odhad předchozí činnosti, 5) zapamatování si oblečení, 6) zapamatování si pozice jednotlivých osob na obrázku a 7) odhad doby, která uplynula od poslední návštěvy. (převzato z Klimeš, 2001)

vnitřní reprezentaci okolního světa, ale že se spoléhá na rychlou dostupnost všech potřebných informací prostřednictvím sákladických očních pohybů, které mozkem požadované informace vyzvedávají ze světa jako ze specifického druhu externího paměťového zařízení, které je snadno dostupné 24 hodin denně (O'Regan, 1992).

To, že lidský mozek využívá této strategie, naznačuje nekumulativní povaha sákladických očních pohybů, která se projevuje tím, že oblast ostrého vidění se často opakovaně vrací na stejná místa pozorovaného výjevu. V případě, že by si mozek vytvářel detailní vnitřní model okolního prostředí, potom lze celkem rozumně předpokládat, že by k těmto opakovaným

návratům nedocházelo, neboť mozek by měl již všechny potřebné informace uložené u sebe. Tuto úvahu potvrzuje výzkum tzv. „slepoty ke změnám“, který ukazuje, že člověk má překvapivě velké problémy s identifikací poměrně velkých změn v pozorovaném výjevu v důsledku toho, že ve své mysli kóduje jen povšechné a obecné informace o tom, co se okolo něj právě nachází a děje. Například psychologové *Daniel Simons* a *Daniel Levin* (1997) provedli výzkum, ve kterém v reálných podmínkách zkoumali, kolik osob si všimne, že v průběhu odpovídání na dotaz studenta na cestu do centra města došlo k záměně osoby tazatele.¹ Zjistili, že polovina náhodných kolemjdoucích si záměny vůbec nevšimla. V jiném výzkumu zase Simons promítal účastníkům experimentu minutový videozáznam basketbalového zápasu a dal jim za úkol spočítat přihrávky mezi členy týmu v bílých dresech. Po pětatřiceti vteřinách vstoupil na scénu figurant v kostýmu gorily, chodil mezi hráči a bušil se do prsou; po devíti vteřinách scénu opustil. Simons zjistil, že jen polovina účastníků experimentu si gorily všimla. (Na přiloženém CD lze v adresáři *Slepota_ke_změnám* nalézt obdobný videozáznam *basketball*, který se od toho původního liší pouze v tom, že mezi hráči se neprochází „gorila“, ale osoba s deštníkem.) Slepota ke změnám se zkoumá také s pomocí tzv. „blikačů“ (*flickers*), kdy se účastníkům experimentu opakovaně, v asi půlvteřinových intervalech promítají dva identické obrázky, které se liší pouze v jednom podstatném detailu. Úkolem osob je tento rozdíl identifikovat. Celý úkol je ztížen tím, že při promítání obrázků je mezi obrázky vložena černá clona, která mozku brání ve využívání pohybových návodů k identifikaci probíhající změny a následně tedy také hledaného rozdílu. K tomuto účelu se kromě černé clony používají také nápadné pohybové distraktory, který odvádějí pozornost člověka od změny signalizující hledaný rozdíl mezi obrázky (něco podobného provádějí kouzelníci, když manipulují pozorností diváků, tak aby jim v určitý kritický okamžik znemožnili získat informace, které by prozradily jejich „kouzlo“). Navzdory tomu, že změny, ke kterým dochází, se často týkají podstatných a výrazných částí obrázku, lidem obvykle trvá překvapivě dlouhou dobu, než se jim podaří rozdíl mezi oběma obrázky identifikovat. (Čtenář si může svůj postřeh sám na několika takových „blikačích“ vyzkoušet - najde je na přiloženém CD v adresáři *Slepota_ke_změnám*.)

¹ V experimentu student s mapou v ruce žádal náhodného kolemjdoucího o radu, jak se dostat co nejrychleji do centra města. Poté, co dotázaná osoba začne studentovi vysvětlovat cestu, přeruší jejich rozhovor dvojice stěhováků nesoucích dveře, kteří mezi nimi projdou, takže dotazovaný studenta na několik vteřin ztratí z dohledu. V této chvíli se student, který se původně ptal na cestu, vyměnil s jiným studentem, který se od něj liší výškou, barvou vlasů, oblečením, tónem hlasu apod. Videonahrávku experimentu je možné shlédnout na <http://viscog.beckman.uiuc.edu/grafts/demos/12.html>. Na přiloženém CD je v adresáři *Slepota_ke_změnám* videozáznam *person_change*, který funguje na základě obdobného principu jako Simonsův a Levinův experiment, jen s tím rozdílem, že namísto dveří je zde k záměně osob využito filmového střihu.

Všechny tyto experimenty potvrzují, že člověk silně přeceňuje množství a především kvalitu informací, které mu zprostředkovává jeho zraková zkušenost. Místo detailního a bohatého vnitřního modelu okolního prostředí si mozek vytváří jen velice hrubý a povšechný obrázek o svém okolí, který je ovšem dostatečně přesný na to, aby ho mohl mozek použít k „navigaci“ sakadických očních pohybů, které mu rychle a podle potřeby dodají všechny informace v dostatečné kvalitě. Mozek namísto detailních informací o okolním světě ukládá spíše jakési meta-informace o tom, jak a kde potřebné informace rychle získat, čímž se značně redukuje zátěž biologické paměti, kterou může z větší části suplovat stabilně přítomné a snadno přístupné vnější prostředí – neboť to, co je na paměti důležité není to, zda se nachází uvnitř, nebo vně organismu, ale to, jak rychle z ní lze požadované informace vyzvednout. Lidský mozek tedy podle všeho využívá stejnou strategii jako Brooksovy roboty: nechává na světě samotném, aby hrál roli svého vlastního modelu a paměti. Iluzorní pocit bohatosti zrakové zkušenosti by pak mohl být dán vědomím, že člověk disponuje snadným a okamžitým přístupem ke všem požadovaným informacím - asi tak jako když má student psychologie subjektivní pocit, že disponuje mnoha detailními znalostmi z obecné psychologie, přestože právě v této chvíli se velká většina těchto poznatků nachází mimo jeho vědomou mysl, on ale ví, že by bylo velice snadné je všechny vědomé mysli zpřístupnit.

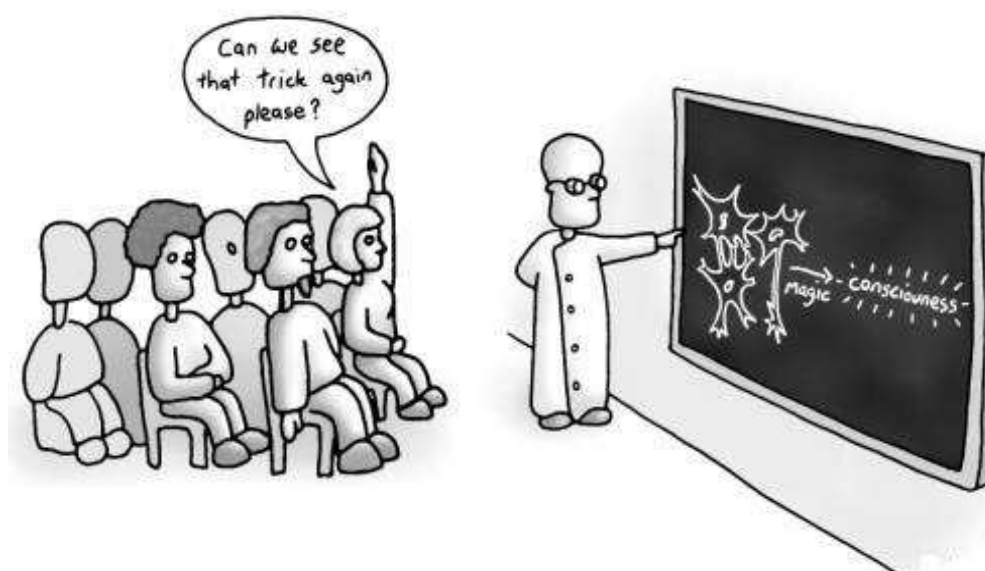
Všechny tyto poznatky zpochybňují základní předpoklad, z něhož vychází tradiční kognitivní věda, totiž že inteligentní systém musí vytvářet detailní vnitřní reprezentaci svého prostředí a následně na této vnitřní reprezentaci realizovat různé výpočetní operace. Ukazuje se, že k využití informací nacházejících se ve vnější prostředí nemusí být tyto informace vnitřně re-reprezentovány v podobě nějaké symbolického kódu¹, ale mohou být prostřednictvím senzorického aparátu využity přímo ke generování a tvarování chování inteligentního systému, který takto tvoří s vnějším prostředím těsně propojený systém², jehož dynamiku lze nejlépe uchopit a studovat prostřednictvím pojmových nástrojů teorie dynamických systémů. Dynamická kognitivní věda takto v jakési nové syntéze poněkud obrušuje ostré hroty dvou jinak nesmiřitelných tezí behaviorismu a kognitivismu.

¹ Reprezentacionalismus, který charakterizuje západní myšlení již někdy od Descarta, je předmětem kritiky také v některých proudech současné filozofie, například v tzv. *postanalytické filozofii* (Peregrin, 1998). Asi nejexplicitněji je tato kritika vyjádřena u *Richarda Rortyho* (2000), který ve své slavné knize *Filozofie a zrcadlo přírody* podrobil důkladné analýze všechny slepé uličky, do kterých vede představa člověka jako „skelné esence“, která vnitřně reprezentuje či odráží vnější svět. Jiný představitel postanalytické filozofie, „*Robert Brandom* cituje vzrušenou reakci *Rebecy Westové*, která nemohla pochopit, proč by měl někdo chtít po lidské mysli, aby vytvářela kopii světa: „Jedna taková zatracená věc přece úplně stačí!““ (Peregrin, 1999, s. 39)

² V případě člověka je potřeba doplnit tento základní obrázek kognitivního systému těsně propojeného se svým prostředím o zásadní vliv, který mají na možnosti lidské kognice nejružnější kulturní artefakty (viz oddíl „1.2.3.5 Distribuované kognitivní systémy“).

2.2.2 Matematické modelování dynamiky šíření a zpracovávání informací v lidském mozku¹

Typickým příkladem dynamického systému je také samotný lidský mozek zpracovávající informace. I tento systém lze uchopit prostřednictvím řady numerických proměnných definujících jeho stav a prostřednictvím matematických rovnic definujících dynamiku chování a vzájemnou interakci těchto proměnných v čase. Stejně jako při vytváření jakéhokoli jiného modelu, i v případě matematického modelování se člověk nevyhne rozhodnutí, které z prvků a charakteristik modelovaného systému do matematického modelu zahrnout a které z něj naopak vypustit. Také matematický model je tedy jistou redukcí či zjednodušením reality a kompromisem mezi reprezentativností a srozumitelností. V případě konekcionistického modelování kognitivních funkcí vychází matematický model (zpracovávání informací v lidském mozku) z pěti principů popsaných v oddíle 2.1. Tyto principy představují (v některých svých aspektech) značně zjednodušený popis výpočetní činnosti lidského mozku na určité specifické úrovni abstrakce, která je podle většiny představitelů konekcionismu tou správnou úrovní analýzy, která umožňuje pochopit, jak lidský mozek realizuje různé mentální stavy a procesy. Na konekcionistické modely dílčích kognitivních funkcí se pak lze dívat jako na určitou formu experimentů, které ověřují, zda a nakolik dokáže dynamika takového masivně paralelně distribuovaného systému realizovat vybranou kognitivní funkci.



Obrázek 120

¹ Obsah celého tohoto oddílu je založen především na pracích následujících autorů: Plunkett, Elman, (1997), McLeod, Plunkett, Rolls (1998), Novák a kol. (1998), Churchland, Sejnowski (1999) a Fanta (2000, 2001).

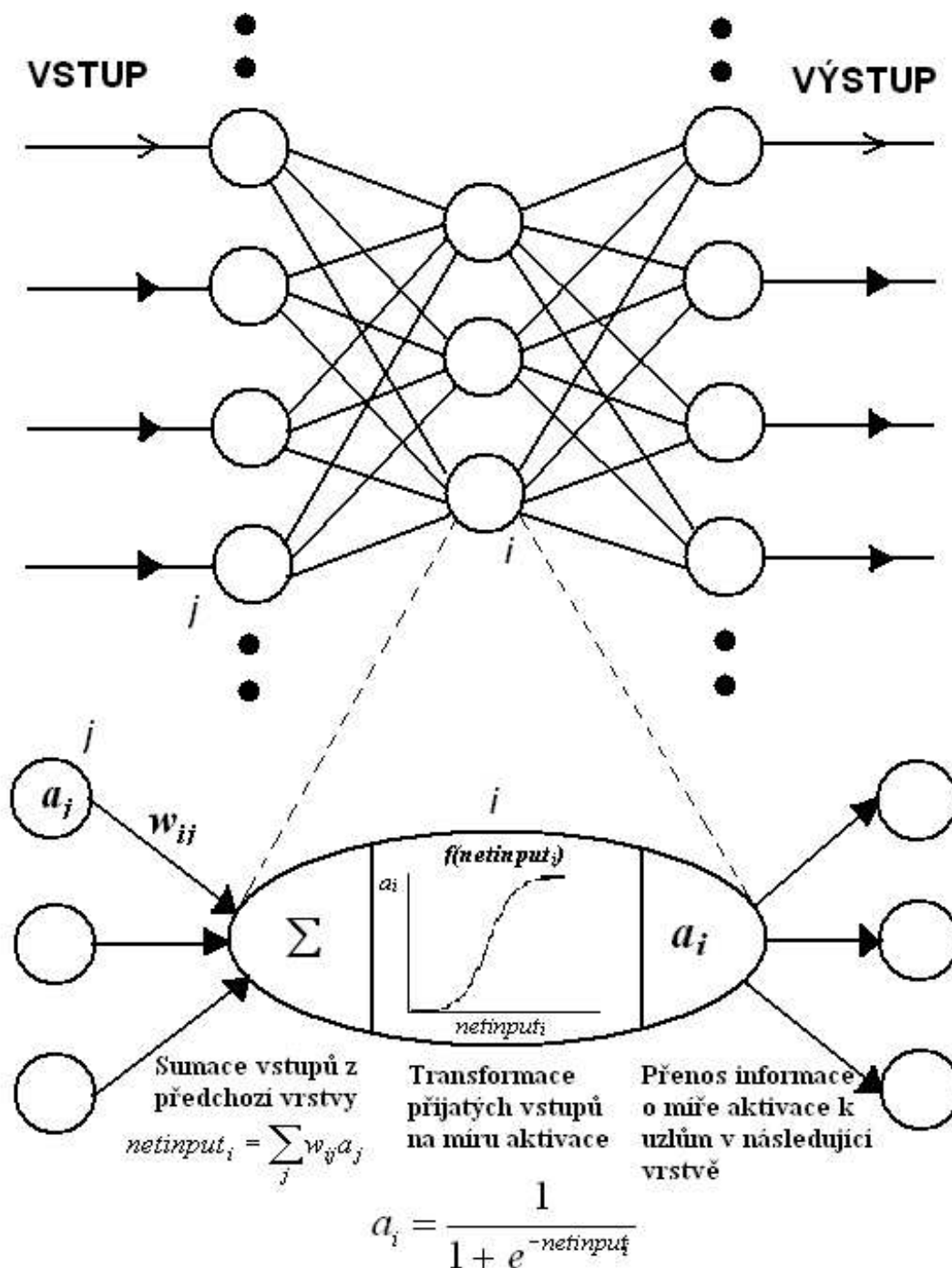
2.2.2.1 Základní pojmy

Umělé neuronové sítě jakožto matematické modely výpočetní činnosti lidského mozku nejsou ničím jiným než souborem různých numerických proměnných a různých matematických funkcí a právě jejich přehledu je věnován tento oddíl. Z oddílu „1.2.3.3 Konekcionismus“ již víme, že konekcionistická síť se skládá ze dvou základních funkčních prvků, a to z **uzlů** a ze **spojů** mezi nimi, které jsou typicky uspořádány v podobě třívrstevné neuronové sítě, která se skládá ze vstupní, skryté a výstupní vrstvy uzlů. Uzly představují jednoduché výpočetní jednotky, jejichž aktuální stav je vždy charakterizován určitou **mírou aktivace**, kterou počítají na základě míry aktivace dalších uzlů nacházejících se v předchozí vrstvě, na základě **síly spojů**, kterými jsou propojeny, a na základě tzv. **aktivační funkce**. Informaci o spočítané míře své aktivace poté uzly odesílají k uzlům nacházejícím se v následující vrstvě. Tímto způsobem se v neuronové síti šíří informace o míře aktivace jednotlivých uzlů směrem od výstupní vrstvy přes vrstvu skrytou k vrstvě vstupní: Neuronové síti je nejdříve prezentován nějaký podnět, který způsobí, že se aktivují některé uzly ve vstupní vrstvě; ty vyvolají určitý vzorec aktivity ve skryté vrstvě, který zase způsobí aktivaci některých uzlů ve vrstvě výstupní, která představuje odpověď neuronové sítě na prezentovaný podnět. Jestliže neuronová síť dokáže v reakci na prezentaci určitého aktivačního vzorce na vstupní vrstvě vyprodukovat ve výstupní vrstvě ten správný vzorec aktivace, potom to znamená, že neuronová síť implementuje danou funkci, resp. že dokáže řešit daný typ problému. Po tomto stručném shrnutí a připomenutí toho, co všechno již víme o principech fungování konekcionistických sítí, se můžeme podrobněji podívat na jednotlivé fáze zpracovávání informací v tomto typu výpočetní architektury a na to, jak je lze matematicky modelovat (McLeod, Plunkett, Rolls, 1998):

1) *Míra aktivace*. Míra aktivace uzlu sítě se označuje symbolem a ; dolní index připojený k tomuto symbolu pak specifikuje konkrétní neuron, jehož míru aktivace symbol označuje; symbolem a_j je takto reprezentována míra aktivace neuronu j . Je přitom zavedeným zvykem, že písmenem j se vždy označuje uzel, který odesílá informaci o své míře aktivace, a písmenem i pak uzel, který tuto informaci přijímá. Míra aktivace uzlu může teoreticky nabývat libovolných hodnot, v praxi se však míra aktivace uzlů neuronových sítí (na základě čistě arbitrárního rozhodnutí) pohybuje většinou mezi 0 a 1.

2) *Váha spojů*. Jednotlivé uzly sítě se ovlivňují prostřednictvím výměny informací o svých mírách aktivace. Síla tohoto vlivu je dána vahou spoje, který existuje mezi vysílajícím a přijímajícím uzlem. Váha spoje se označuje symbolem w_{ij} , kde první dolní index označuje

přijímající uzel a druhý pak uzel vysílající. Čím větší je hodnota váhy spoje, tím má vysílající uzel větší vliv na míru aktivace uzlu přijímajícího. Váha spoje přitom může být buď pozitivní, nebo negativní. V případě, že je váha spoje pozitivní, vysílající uzel míru aktivace uzlu přijímajícího zvyšuje; v případě negativní váhy spoje vysílající uzel míru aktivace uzlu přijímajícího naopak snižuje (na rozdíl od biologických neuronů, jejichž vliv na okolní



Obrázek 121: Výpočetní operace základní jednotky konekcionistického systému. Výpočetní činnost uzlů neuronové sítě spočívá v jednoduché sumaci vstupů z předchozí vrstvy a v aplikaci aktivační (přechodové) funkce na tento součet. Výsledkem těchto výpočtů je míra vlastní aktivace, kterou uzel odesílá uzlům v následující vrstvě. (s úpravami převzato z McLeod, Plunkett, Rolls, 1998, s. 16)

neurony může být buď pouze excitační, nebo pouze inhibiční, výpočetní jednotky umělých neuronových sítí jsou obvykle zároveň excitační i inhibiční; proto jsou také uzly umělých neuronových sítí považovány spíše za model skupiny neuronů než jednoho neuronu). Váha spoje může nabývat prakticky libovolných kladných nebo záporných hodnot.

3) *Vstup*. Vstup ($input_i$) do uzlu i z uzlu j je funkcí míry aktivace uzlu j a váhy spoje vedoucího z uzlu j do uzlu i (w_{ij}):

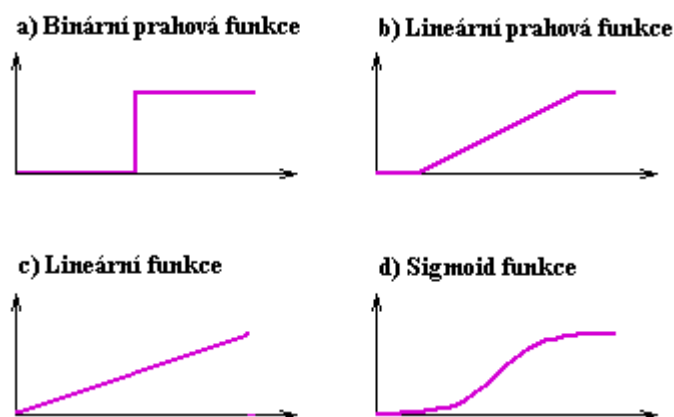
$$input_{ij} = a_j w_{ij}$$

4) *Celkový vstup*. Celkový vstup do uzlu i ($netinput_i$) je sumou vstupů ze všech uzlů, od kterých uzel i prostřednictvím vážených spojů přijímá informace o jejich míře aktivace:

$$netinput_i = \sum_j w_{ij} a_j$$

Celkový vstup do uzlu může opět nabývat libovolných kladných či záporných hodnot.

4) *Aktivační funkce*. Aktivační (také přechodová) funkce transformuje celkový vstup do uzlu na jeho míru aktivace. Nejjednodušší aktivační funkce má podobu *funkce identity*, kdy se míra aktivace jednoduše rovná celkovému vstupu do uzlu. Přehled nejpoužívanějších typů aktivačních funkcí je uveden na obrázku 122: a) Binární prahová funkce dovoluje uzlu sítě nabývat pouze dvou možných hodnot (1/0), kdy je daný uzel buď aktivní, nebo neaktivní; uzel se přitom může stát aktivní pouze tehdy, když vstup překročí určitou kritickou (prahovou) hodnotu θ . b) Lineární prahová funkce rovněž vyžaduje určitou minimální míru celkového vstupu než dojde k aktivaci uzlu sítě; na rozdíl od binární prahové funkce však umožňuje uzlu nabývat libovolných hodnot míry aktivace od 0 do ∞ v (lineární) závislosti na velikosti celkového vstupu. c) Lineární funkce je obdobou lineární prahové funkce s tím

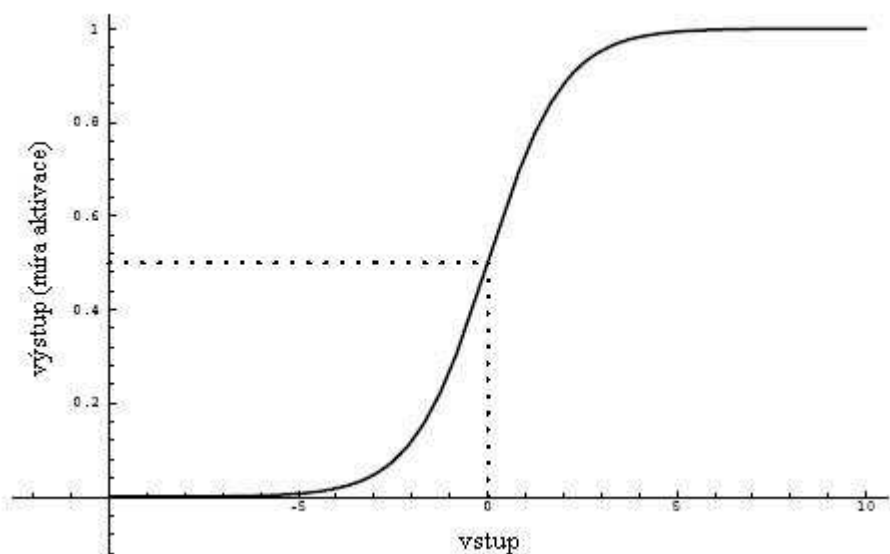


Obrázek 122: Možné typy aktivačních funkcí transformujících celkový vstup do uzlu ($netinput_i$) na míru jeho aktivace (a_i).

jediným rozdílem, že k aktivaci uzlu zde není potřeba překonávat žádnou minimální (prahovou) míru celkového vstupu do uzlu. d) V konekcionistickém modelování se nejčastěji používá nelineární sigmoid aktivační funkce, která představuje biologicky plausibilní model chování skutečných neuronů, jejichž míra aktivity se pohybuje jen v určitém omezeném rozmezí hodnot, takže narůstající nebo naopak klesající hodnota vstupu nemá na míru aktivace neuronu příliš velký vliv, jestliže se tento již nachází v blízkosti své maximální, resp. minimální možné míry aktivace. Konkrétním příkladem nelineární sigmoid aktivační funkce je logistická aktivační funkce, kterou vyjadřuje rovnice:

$$a_i = \frac{1}{1 + e^{-netinput_i}} ,$$

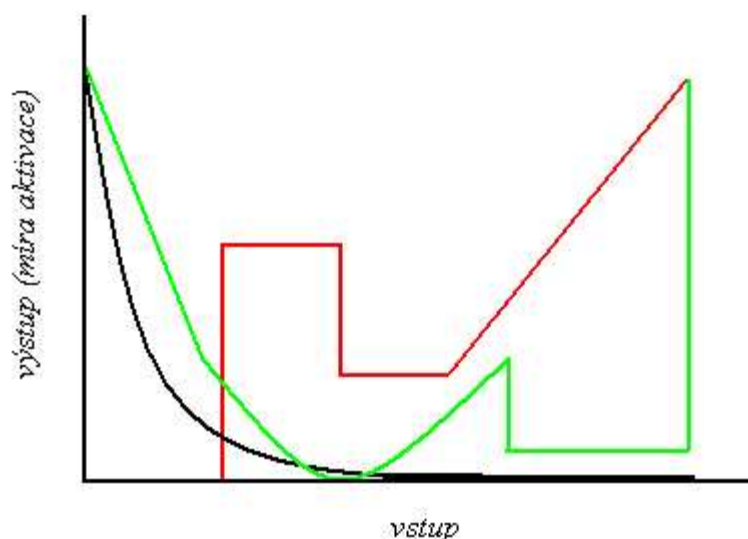
kde a_i označuje výslednou míru aktivace, $netinput_i$ celkový vstup do uzlu a e exponenciál. Na obrázku 123 je zobrazen průběh této funkce pro $e = 2,72$: Pro velkou část hodnot celkového vstupu (-4 a méně, resp. 4 a více) tato aktivační funkce vykazuje chování ve stylu „vše nebo nic“, kdy se odpovídající míra aktivace pohybuje blízko nulové, resp. maximální možné hodnotě ¹; v případě takto extrémních hodnot celkového vstupu má jejich jakýkoli další pokles, resp. nárůst jen minimální vliv na výslednou míru aktivace; jinými slovy, jakýkoli nárůst či pokles celkového vstupu je zde v nepoměru k odpovídající změně v míře aktivace daného uzlu. V rozmezí hodnot celkového vstupu od -4 do 4 však logistická aktivační funkce umožňuje relativně citlivé rozlišování mezi různými úrovněmi vstupů do uzlu: Jak se míra celkového vstupu stává stále méně negativní, odpovídající míra aktivace narůstá, nejdříve



Obrázek 123: Logistická aktivační funkce.

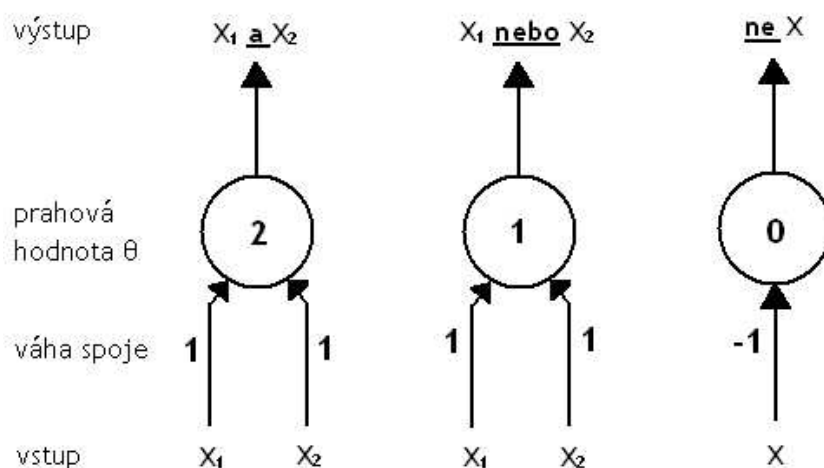
¹ Při použití logistické aktivační funkce míra aktivace daného uzlu nikdy nedosáhne nulové, ani maximální možné hodnoty (1); míra aktivace uzlu se těmito krajními hodnotám může pouze nekonečně přibližovat.

pozvolna a potom stále rychleji; s nárůstem kladných hodnot celkového vstupu se pak nárůst odpovídající míry aktivace opět postupně zpomaluje a velice pozvolna se přibližuje k maximální možné hodnotě. Vedle těchto čtyř nejčastěji používaných aktivačních funkcí je teoreticky možné uvažovat o použití téměř nekonečné množství dalších aktivačních funkcí, které mohou vyjadřovat i velice neobvyklé vztahy závislosti mezi mírou aktivace uzlu a hodnotami jeho celkového vstupu (viz obrázek 124).



Obrázek 124: Další možné podoby aktivační funkce.

V roce 1943 ve své práci *A Logical Calculus of the Ideas Immanent in Nervous activity* (Logický kalkulus myšlenek imanentních v nervové činnosti) W. McCulloch a Walter Pitts prokázali, že s pomocí prahového zařízení (uzlu s binární prahovou aktivační funkcí), které je zjednodušeným matematickým modelem činnosti biologického neuronu, lze velice snadno implementovat základní logické jednotky „a“, „nebo“ a „invertor“ (tzv. logická hradla, viz s. 60-62): Logickou jednotku „a“, která na výstupu vyprodukuje 1 pouze tehdy, když jsou oba vstupy 1, lze realizovat prostřednictvím prahového zařízení, jehož dva vstupy mají váhu 1 a jehož prahová hodnota θ se rovná 2; k implementaci logické jednotky „nebo“, která na výstupu poskytuje 1 tehdy, pokud alespoň jeden vstup je 1, postačí změnit nastavení prahové hodnoty θ na 1; a v případě logické jednotky „invertor“, která mění vstup 1 na výstup 0 a vstup 0 na výstup 1, se prahová hodnota θ nastaví na 0 a vstupu se přisoudí váha -1 (viz obrázek 125). Lze říci, že to, co v roce 1940 udělal Shannon pro počítač a pro jeho relátka (viz s. 60-62), udělali jen o tři roky později McCulloch a Pitts pro mozek a pro jeho neurony: Ukázali, že správným zapojením těchto jednoduchých výpočetních zařízení je teoreticky možné implementovat jakoukoli binární logickou funkci, kterou lze zapsat jako přesný



Obrázek 125: Základní logické jednotky „a“, „nebo“ a „inverze“ implementované v podobě jednoduchých prahových zařízení (uzlů umělé neuronové sítě s binární prahovou funkcí).

logický výrok. To jinými slovy znamená,

že umělá neuronová síť je univerzálním výpočetním zařízením (*universal function approximator*), které dokáže teoreticky realizovat jakoukoli funkci, kterou dovede spočítat Turingův stroj, resp. digitální počítač, který je jeho hmotnou inkarnací. Rumelhart a kol. (1986b) v této souvislosti konstatovali, že „*kdybychom měli ty správné spoje vedoucí ze vstupních jednotek do dostatečně velkého souboru skrytých jednotek, potom bychom vždy dokázali najít reprezentaci, která by zajistila jakékoli mapování vstupů na výstupy*“¹. Avšak přestože je teoreticky možné sestavit neuronovou síť schopnou realizovat jakýkoli dostatečně dobře specifikovaný výpočet, nic člověku nezaručuje, že se mu v pomyslném prostoru všech možných neuronových sítí podaří v rozumném čase najít právě takovou neuronovou síť, která bude vhodná k realizaci daného výpočtu. Jinak řečeno, „*teoreticky může být jakákoli úloha vyřešena nějakou neuronovou sítí, ale ne každá neuronová síť dokáže vyřešit jakoukoli úlohu*“² (Plunkett, Elman, 1997, s. 27). Je to jako vědět, že v pytli plném padesátníků se někde zcela určitě nachází padesátník, který jsem včera použil k placení piva v hospodě „U Nováků“: Jedna věc je ovšem vědět, že se můj padesátník v pytli plném dalších padesátníků určitě nachází, a druhá věc je být schopen tento konkrétní padesátník v pytli skutečně najít. Onen pomyslný prostor všech možných neuronových sítí je definován zejména nastavením váhové matice a architekturou neuronové sítě, tj. počtem a způsobem zapojení výpočetních jednotek. Vytváření konekcionistického modelu nějaké kognitivní funkce tak tedy spočívá v prohledávání tohoto prostoru a v hledání neuronové sítě s takovou váhovou maticí a s takovou architekturou, která bude vykazovat požadované chování (tím či oním způsobem srovnatelné s chováním pozorovaným u lidských subjektů).

¹ “If we have the right connections from the input units to a large enough set of hidden units, we can always find a representation that will form any mapping from input to output.”

² „... in theory, any task can be solved by some neural network, not any neural network can solve any task.“

2.2.2.2 Automatické učící algoritmy

Jestliže je míra aktivace uzlu i (a_i) funkcí sumy všech jeho vstupů ($netinput_i$) a jestliže je každý z těchto vstupů ($input_i$) funkcí míry aktivace uzlů, které jsou s uzlem i spojeny, násobené vahou spojení mezi nimi ($a_j w_{ij}$), potom je zřejmé, že má-li být umělá neuronová síť schopna požadované transformace vstupního vzorce aktivace na výstupní aktivační vzorec, musí disponovat určitým specifickým souborem vah, který jí tuto transformaci umožní realizovat. Soubor všech vah umělé neuronové sítě (dále UNS) tvoří tzv. váhovou matici W , která je definována jako matice $n \times n$, kde w_{ij} je váha synaptického spoje mezi uzly i a j .

$$W = \begin{bmatrix} w_{11} & w_{12} & \dots & \dots \\ w_{21} & w_{22} & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & w_{nn} \end{bmatrix}, \text{ kde } w_{ii} = 0$$

Tato váhová matice definuje $n \times n$ -rozměrný prostor, který představuje množinu všech možných váhových matic pro danou UNS. Hledání souboru vah umožňujícího realizaci požadované funkce je pak prohledáváním tohoto vícedimenzionálního prostoru a hledáním souřadnic (číselných hodnot jednotlivých vah) lokalizujících váhovou matici kódující „mentální“ procesy, které by zajistily požadovanou transformaci vstupů na výstupy. Toto prohledávání váhového prostoru UNS je obdobou procesu řešení problému chápaného Newellem a Simonem jako prohledávání symbolického stavového prostoru daného problému a hledání stavu odpovídajícího jeho řešení (viz s. 74-75). A podobně jako lidé k usnadnění „navigace“ rozsáhlým problémovým prostorem používají různé heuristiky (viz s. 76-77), tak i při prohledávání váhového prostoru UNS se používá řada automatických procedur, tzv. **učících algoritmů** či **pravidel** (*learning algorithms/rules*), která umožňují se vyhnout jinak prakticky nezvládnutelné úloze ručního nastavení vah všech spojů v UNS.

Tyto automatické učící algoritmy stojí v pozadí schopnosti UNS se učit, tj. měnit svoje chování na základě svých vlastních zkušeností. Na rozdíl od běžných počítačových modelů nejsou UNS programovány – alespoň ne v tradičním slova smyslu, kdy programátor musí specifikovat všechna pravidla týkající se způsobu řešení daného souboru úloh. V případě UNS programátor pouze nastaví nějaké základní parametry sítě (počet uzlů, jejich konektivita, použitá aktivační funkce), učícího algoritmu a podnětového materiálu (tzv. *tréninkového prostředí*) a nechá již na samotné UNS, aby na základě statistické struktury

implicitně obsažené v poskytnutém vzorku příkladů vstupně-výstupních dvojic induktivně objevila nějaké obecné pravidlo či zákonitost, kterou lze generalizovat na celou množinu podobných úloh. UNS takto v pozadí podnětového materiálu často dokáže objevit i velice jemné a abstraktní zákonitosti - některé UNS de facto provádí určitý druh shlukové analýzy, multidimenzionálního škálování či analýzy hlavních komponent (Fanta, 2000).

2.2.2.2.1 Hebbovo pravidlo

Jedním z nejpoužívanějších učících algoritmů je známé Hebbovo pravidlo (*Hebb rule* nebo také *hebbian learning rule*), které navrhl kanadský neurovědec Donald Hebb jako možný biologický mechanismus jednoduchého asociačního učení, které spočívá ve vytváření jednoduchých asociačních spojů mezi různými podněty. Jedním takovým podnětem by mohla být například rozsvícená lampa a druhým podnětem pak miska plná dobrého jídla – použijeme-li známý příklad z Pavlovových pokusů s jednoduchým podmiňováním u psů. Tyto dva podněty v mozku psa vyvolají dva různé a na sobě nezávislé vzorce nervové aktivity, které plní funkci jejich mentálních reprezentací. Mezi dvěma podněty pro psa původně neexistuje žádné spojení; jinými slovy, nijak spolu podle něj nesouvisí. Avšak díky tomu, že jsou psovi oba podněty opakovaně prezentovány zároveň nebo těsně za sebou, mezi odpovídajícími vzorci nervové aktivity se postupně vytvoří asociační spoj, takže až pes příště uvidí svítící lampu, může to v něm vyvolat reakci původně spojenou pouze se zahlednutím misky plné jídla (například zvýšenou salivaci). Asociační učení takto umožňuje získat ekologicky významné informace o kauzální struktuře světa, o tom, jak spolu různé věci souvisí, jaké jsou jejich vlastnosti a co od nich lze očekávat.¹ Při osvojování si asociačního spoje mezi dvěma podněty je klíčová souběžná prezentace obou podnětů, neboť právě ta stojí v samotných základech Hebbova pravidla, které spočívá v zesilování synaptické váhy mezi těmi neurony/uzly, které jsou současně aktivní. Ve formální notaci má Hebbovo pravidlo podobu následujícího algebraického výrazu

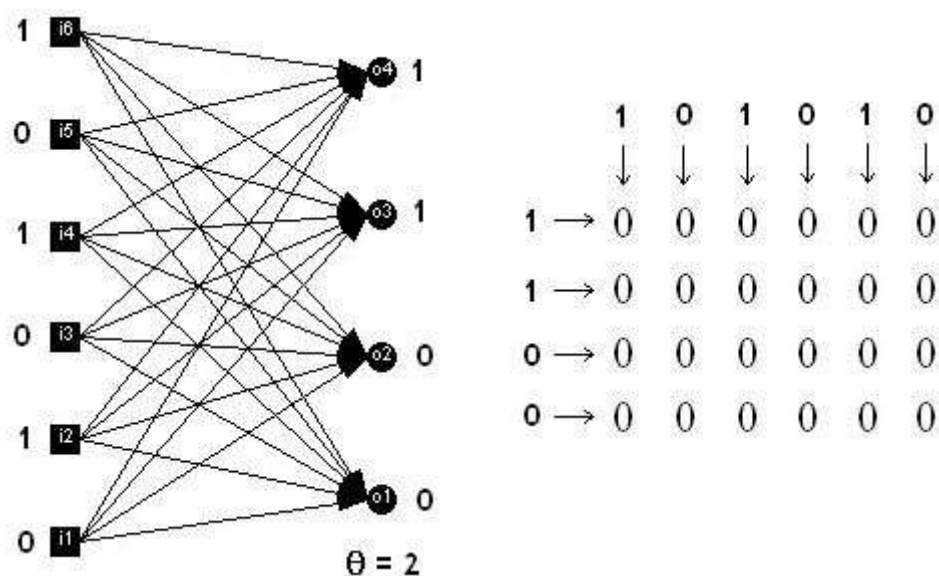
$$\Delta w_{ij} = \varepsilon a_i a_j ,$$

kde Δw_{ij} je změna váhy spoje mezi uzly i a j , ε je míra učení (*learning rate*) – konstanta specifikující velikost korekce váženého spoje v jednom kole učení (*learning trial*), a_i je míra

¹ „Náš mozek... automaticky zaznamenává dvě základní stránky naší zkušenosti, a to (1) souvislosti se přirozeně vyskytujícími mezi jednotlivými podněty a (2) vztah mezi chováním a jeho důsledky. Z těchto dvou je schopnost automaticky odhalovat souvislosti mezi událostem základnější.“ (Cumminsová, 1998, s. 117)

aktivace uzlu i a a_j je míra aktivace uzlu j . Násobení míry aktivace dvou uzlů je vyjádřením základní myšlenky Hebbova pravidla, totiž že k posílení váhy spoje mezi dvěma uzly může dojít pouze tehdy, když jsou oba uzly současně aktivní.

Aplikaci Hebbova pravidla můžeme demonstrovat na jednovrstevné neuronové síti (nazývané také *perceptron* nebo také *asociátor vzorců*), jejímž úkolem bude osvojit si asociaci mezi rozsvícenou lampou – reprezentovanou ve vstupní vrstvě aktivačním vektorem $[1\ 0\ 1\ 0\ 1\ 0]$ – a miskou plnou jídla – reprezentovanou ve výstupní vrstvě aktivačním vektorem $[1\ 1\ 0\ 0]$.¹ Na obrázku 126 je schematicky znázorněna UNS a její váhová matice: Jednotlivé uzly používají binární prahovou aktivační funkci a práh θ má hodnotu 2; každý řádek váhové matice reprezentuje váhy spojů z šesti uzlů vstupní vrstvy do jednoho uzlu vrstvy výstupní; před začátkem učení jsou váhy všech spojů nastaveny na 0. Aplikací

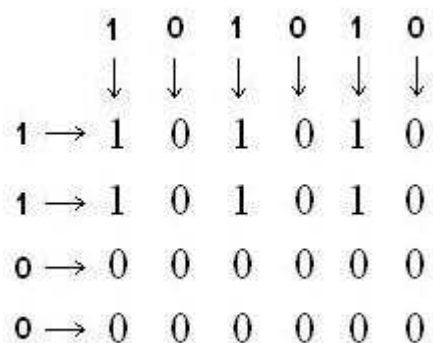


Obrázek 126: Nalevo je jednovrstevná neuronová síť, jejímž úkolem je se naučit asociovat rozsvícenu lampu - reprezentovanou ve vstupní vrstvě aktivačním vektorem $[1\ 0\ 1\ 0\ 1\ 0]$ - s miskou plnou jídla – reprezentovanou ve výstupní vrstvě aktivačním vektorem $[1\ 1\ 0\ 0]$. Napravo je odpovídající váhová matice ve stavu před začátkem procesu učení, kdy jsou všechny vážené spoje nastaveny na 0. (s úpravami převzato z McLeod, Plunkett, Rolls, 1998, s. 55)

Hebbova pravidla s $\varepsilon = 1$ v rámci jednoho kola učení se váhy změní do podoby, která je zachycena ve váhové matici na obrázku 127, kde je dobře vidět, že váhy se zesílily pouze mezi těmi uzly, které byly současně aktivní, tzn. spoje vedoucí z prvního, třetího a pátého uzlu vstupní vrstvy do prvního a druhého uzlu vrstvy výstupní. To jestli si UNS dokázala

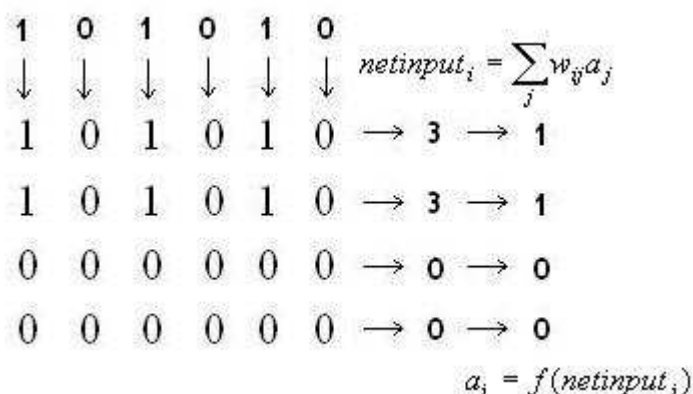
¹ Uvedený příklad je založen na ukázce použití Hebbova pravidla, kterou ve své práci použili McLeod, Plunkett a Rolls (1998, s. 54-58).

osvojit asociaci mezi rozsvícenou lampou a miskou jídla lze jednoduše zjistit tak, že se UNS ve vstupní vrstvě předloží aktivační vektor reprezentující rozsvícenou lampu a poté se sleduje, zda si UNS ve výstupní vrstvě vybaví správný vzorec aktivace, který reprezentuje misku jídla. Na obrázku 128 je zachycen výpočetní proces spojený se zpracováním vstupního vzorce



Obrázek 127: Váhová matice po jednom kole učení s pomocí Hebbova pravidla. (s úpravami převzato z McLeod, Plunkett, Rolls, 1998, s. 55)

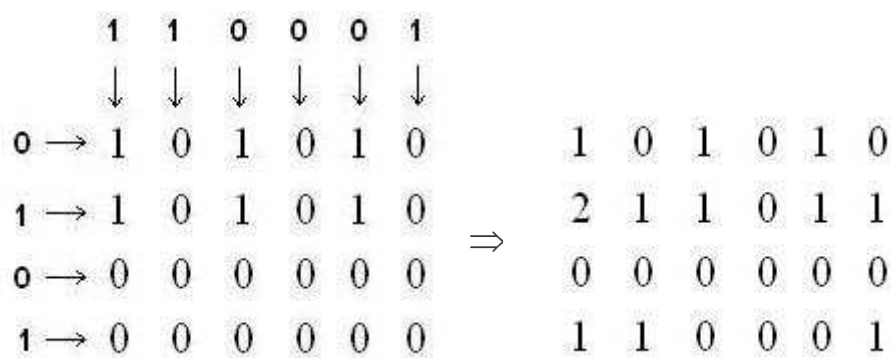
aktivace: Celkový vstup ($netinput_i$) do prvních dvou uzlů výstupní vrstvy má hodnotu 3 a do zbývajících dvou uzlů je roven 0; při použití binární prahové aktivační funkce s prahovou hodnotou $\theta = 2$ to znamená, že výsledný vzorec aktivace má podobu [1 1 0 0]. UNS se tedy úspěšně podařilo osvojit si asociaci mezi rozsvícenou lampou a miskou jídla.



Obrázek 128: Proces asociace aktivačního vzorce [1 1 0 0] na základě prezentace vstupního vzorce aktivace [1 0 1 0 1 0]. (s úpravami převzato z McLeod, Plunkett, Rolls, 1998, s. 56)

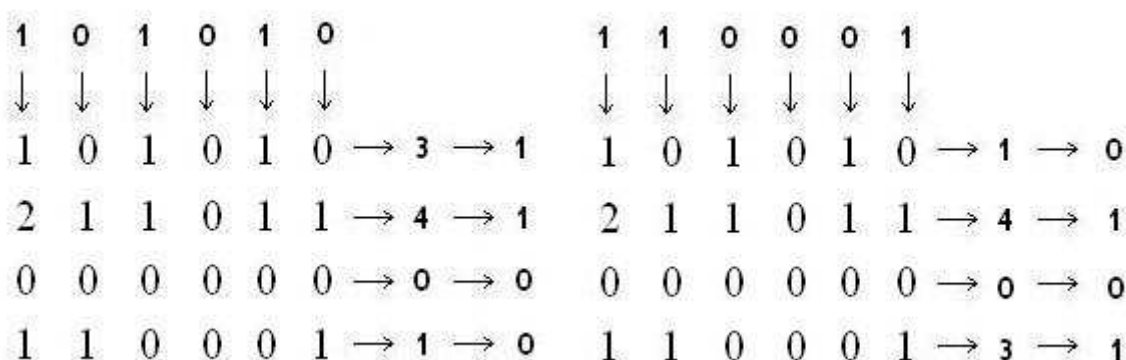
Do stejné váhové matice lze zakódovat i další asociace, například mezi zvukem klaksonu – reprezentovaným na vstupu aktivačním vektorem [1 1 0 0 0 1] - a elektrickým šokem – reprezentovaným ve výstupní vrstvě aktivačním vektorem [0 1 0 1]. V levé části obrázku 129 je stav váhové matice UNS před započítáním nového kola učení s novou dvojicí podnětů; v pravé části je pak výsledná váhová matice, která je de facto kombinací dvou váhových matic

získaných při asociaci dvou různých párů podnětů (rozsvícená lampa/miska jídla, zvuk klaksonu/elektrický šok). Na obrázku 130 je pak vidět, že UNS s touto váhovou maticí si při prezentaci rozsvícené lampy, resp. zvuku klaksonu vždy dokáže na výstupní vrstvě správně vybavit misku jídla, resp. elektrický šok a to přesto, že v průběhu procesu asociace došlo



Obrázek 129: Nalevo je váhová matice ve stavu před dalším kolem učení, kdy si má neuronová síť osvojit asociaci mezi zvukem klaksonu [1 1 0 0 0 1] a elektrickým šokem [0 1 0 1]. Napravo je pak výsledná váhová matice, ve které jsou zakódovány obě asociace, jak mezi rozsvícenou lampou a miskou jídla, tak i mezi zvukem klaksonu a elektrickým šokem. (s úpravami převzato z McLeod, Plunkett, Rolls, 1998, s. 57)

k určité interferenci mezi váhovými maticemi kódujícími jednotlivé asociace. Tuto interferenci (tzv. „crosstalk“) se zde podařilo úspěšně vykompenzovat pomocí dostatečně vysoké prahové hodnoty. Přesto je zřejmé, že zde musí existovat jisté omezení v množství asociací, které je možné do této konkrétní váhové matice zakódovat, aniž by přitom docházelo ke katastrofickým selháním ve vybavování jednotlivých asociací. Jakkoli je existence interference mezi podobnými reprezentačními kódy nevýhodou v případě vytváření nějaké



Obrázek 130: Ověření spolehlivosti procesu vybavování dvou různých asociací, které jsou zakódovány do jedné váhové matice. Nalevo je zachycen proces vybavování misky jídla [1 1 0 0] při prezentaci rozsvícené lampy [1 0 1 0 1 0], napravo je pak zaznamenán proces asociace zvuku klaksonu [1 1 0 0 0 1] a elektrického šoku [0 1 0 1]. (s úpravami převzato z McLeod, Plunkett, Rolls, 1998, s. 58)

komerční automatizované databáze, v případě výpočetního modelování kognitivních funkcí se jedná o psychologicky velice plausibilní charakteristiku procesu vybavování, kterou lze pozorovat i u člověka. Například ve snaze zapamatovat si nové telefonní číslo, které je podobné nějakému jinému známému telefonnímu číslu, se člověku může snadno stát, že si číslice obou telefonních čísel poplete a nakonec si není schopen správně vybavit ani jedno telefonní číslo.

2.2.2.2.2 Delta pravidlo

Dalším často používaným učícím algoritmem je delta pravidlo (*delta learning rule*). Základem tohoto pravidla je výpočet chyby δ_i na uzlu i ve výstupní vrstvě, kterou lze velice snadno zjistit odečtením skutečné míry aktivace uzlu i (a_i), která je výsledkem působení excitačních a inhibičních spojů z uzlů předchozí vrstvy, od požadované míry aktivace uzlu i (t_i):

$$\delta_i = t_i - a_i$$

Je-li míra aktivace uzlu i příliš nízká, potom je potřeba zesílit váhy spojů z těch uzlů předchozí vrstvy, které do uzlu i vysílají excitační signál, a naopak zeslabit váhy spojů z těch uzlů předchozí vrstvy, které do uzlu i vysílají inhibiční signál. Je-li míra aktivace uzlu i naopak příliš vysoká, korekce vah spojů bude mít přesně opačný směr než v předchozím případě. Ve formální notaci má delta pravidlo podobu následujícího algebraického výrazu

$$\Delta w_{ij} = \varepsilon \delta_i a_j,$$

kde Δw_{ij} je změna váhy spoje mezi uzly i a j , ε je míra učení, která specifikuje velikost korekce váženého spoje v jednom kole učení, δ_i je rozdíl mezi požadovanou a skutečnou hodnotou míry aktivace uzlu i a a_j je míra aktivace uzlu j . Z rovnice je dobře patrné, že v případě příliš vysoké míry aktivace uzlu i , kdy δ_i nabývá záporných hodnot, bude korekce síly váženého spoje (Δw_{ij}) rovněž záporná, a vice versa; a bude-li se δ_i rovnat nule, potom také úprava vah spojů bude nulová. Důležitou roli v rovnici hraje míra aktivace vysílajícího uzlu j (a_j), která umožňuje zohlednit „míru spoluviny“ daného uzlu za chybu δ_i při úpravě vážených spojů: Čím je daný uzel aktivnější, tím více se podílí na chybě δ_i , a proto také musí být v odpovídající míře více upravena váha spoje vedoucího z tohoto uzlu. Opakovanou aplikací

delta pravidla se chyba δ bude v závislosti na nastavené míře učení určitým tempem postupně snižovat až dosáhne nulové (nebo nějaké jiné předem stanovené) hodnoty, kdy se proces učení zastaví.

2.2.2.2.3 Perceptronové konvergenční pravidlo

Perceptronové konvergenční pravidlo (*perceptron convergence rule*, dále PKP) je obdobou delta pravidla. Jediný rozdíl mezi nimi spočívá v tom, že kromě úpravy vah spojů podle rovnice

$$\Delta w_{ij} = \varepsilon \delta_i a_j$$

PKP umožňuje rovněž úpravu prahové hodnoty θ podle rovnice

$$\Delta \theta = -\varepsilon \delta_i,$$

kde $\Delta \theta$ je změna prahové hodnoty, ε je míra učení a δ je rozdíl mezi požadovanou a skutečnou hodnotou míry aktivace uzlu i . Vztah mezi chybou δ a korekcí prahové hodnoty ($\Delta \theta$) je zde přesně opačný než v případě chyby δ a úpravy vah spojů (Δw_{ij}): Je-li skutečná míra aktivace uzlu i vyšší než požadovaná míra aktivace, tzn. jestliže je chyba δ záporná, potom má úprava prahové hodnoty kladnou hodnotu (v důsledku dvou znamének „mínus“ v rovnici) a vice versa. Díky tomu se při příliš vysoké míře aktivace uzlu i ztíží možnost překročení kritické prahové hodnoty a následné aktivace uzlu i , a v případě příliš nízké míry aktivace se tato možnost naopak usnadní. Výsledkem opakované aplikace PKP je opět postupná redukce chyby δ . Při použití PKP máme vždy jistotu, že najde hledanou konfiguraci váhové matice a prahových hodnot umožňující realizaci požadované funkce (implementovatelné jednovrstevnou neuronovou sítí), ovšem pouze za předpokladu, že taková konfigurace skutečně existuje.

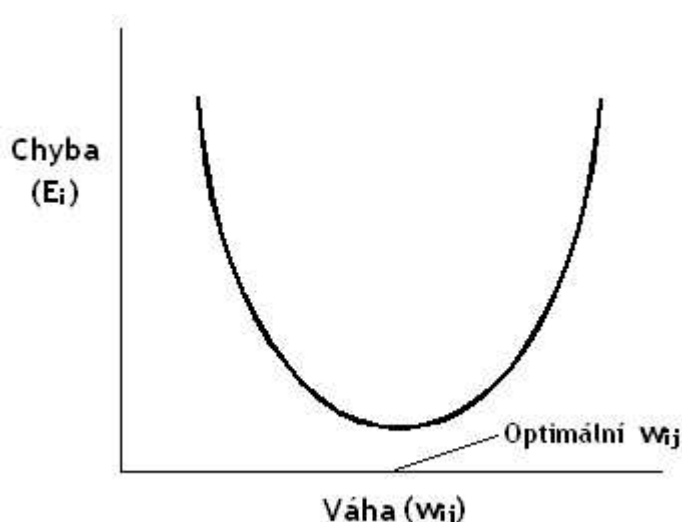
2.2.2.2.4 Pravidlo nejmenších čtverců

Poněkud odlišný přístup k prohledávání váhového prostoru neuronové sítě představuje pravidlo nejmenších čtverců (*least-mean-squares learning procedure*, dále LMS). I tento učící algoritmus vychází z chybového údaje o rozdílu mezi požadovanou a skutečnou mírou

aktivace uzlu i . Na rozdíl od chybových údajů používaných předchozími učícími algoritmy má tento podobu druhé mocniny zjištěného rozdílu, tzn. že chybový údaj bude mít vždy kladnou hodnotu:

$$E_i = (t_i - a_i)^2$$

Stejně jako v předchozích případech, i zde bude chybový údaj E_i variovat v závislosti na hodnotách vah spojů spojujících uzel i s uzly ve vstupní vrstvě. Kdybychom zkonstantnili váhy všech spojů kromě jednoho a kdybychom zakreslili vývoj chybové křivky v závislosti na hodnotě tohoto jednoho váženého spoje, získali bychom chybovou křivku s jedním nejnižším bodem, který odpovídá nejmenší možné chybě při dané hodnotě váhy w_{ij} (viz obrázek 131).



Obrázek 131: Vztah chyby E_i a hodnoty váženého spoje w_{ij} . (s úpravami převzato z McLeod, Plunkett, Rolls, 1998, s. 100)

Cílem učícího algoritmu je pak najít takovou hodnotu váhy w_{ij} , která odpovídá nejnižšímu bodu na chybové křivce. LMS k tomu využívá údaj o sklonu chybové křivky, protože v případě, že známe její sklon, potom lze velice snadno vypočítat, jakým směrem (+/-) je potřeba změnit váhu w_{ij} , aby se chyba E_i zmenšila: Jestliže je sklon negativní, potom to znamená, že w_{ij} je nižší než je potřeba, takže hodnota w_{ij} se zvýší; jestliže je sklon chybové křivky naopak pozitivní, potom to znamená, že w_{ij} je větší než je potřeba a hodnota w_{ij} se sníží. Opakovanou aplikací tohoto pravidla se dříve či později dostaneme na dno chybové křivky, kde je sklon křivky nulový, v důsledku čehož se proces učení zastaví.

Celou proceduru lze připodobnit k situaci, kdy jako horolezci stojíme na svahu velkého kráteru, nevidíme ani na krok před sebe, protože je černočerná tma, a naším úkolem je se

dostat na dno kráteru. Protože nevidíme, nemůžeme se podívat a jít rovnou směrem k cíli. Naší jedinou možností je udělat malý krok a na základě údajů z vnitřního ucha nebo proprioreceptorů zjistit, zda jdeme směrem nahoru, nebo dolů. Jestliže jdeme směrem nahoru, vrátíme se na původní místo a zkusíme udělat krok opačným směrem. Jestliže zjistíme, že tentokrát jsme udělali krok směrem dolů, zůstaneme tam, kde jsme, a zkusíme udělat další krok. Opakováním této procedury skládající se z malých kroků a jednoduché lokální zpětné vazby se budeme pomalu sunout ze svahu kráteru směrem dolů až se nakonec zastavíme na jeho samém dně, kde již nelze udělat žádný krok, který by vedl ještě níže. V této metafoře LMS dno kráteru odpovídá minimální hodnotě chyby E_i , malý krok horolezce odpovídá nepatrnému zvýšení nebo snížení hodnoty váhy w_{ij} a jednoduchá zpětná vazba odpovídá informaci o tom, zda zvýšení nebo snížení váhy w_{ij} vede ke zlepšení nebo naopak ke zhoršení výkonu, tedy zda došlo ke snížení, nebo ke zvýšení chyby E_i (za předpokladu, že všechny ostatní váhy mají konstantní hodnotu). Tato „horolezecká“ povaha LMS se také odráží v jeho dalším názvu, který v originále zní *gradient descent learning*, tedy učení klesáním po svahu; možný je také „zrcadlově převrácený“ název *hill climbing*, doslova šplhání na kopec, častější je však překlad vyhledávání extrému.

Ve formální notaci má LMS podobu rovnice

$$\Delta w_{ij} = -\varepsilon \frac{dE_i}{dw_{ij}},$$

kde Δw_{ij} je opět změna váhy spoje mezi uzly i a j , ε je míra učení specifikující velikost korekce váženého spoje z uzlu j a dE_i/dw_{ij} je derivace chyby E_i podle váhy w_{ij} , která vyjadřuje míru změny E_i ve vztahu k w_{ij} , tj. sklon chybové křivky E_i (viz také poznámka pod čarou 1 na straně 154). Znaménko „mínus“ v rovnici zajišťuje, že při negativním sklonu chybové křivky se hodnota váhy zvyšuje a při pozitivním sklonu snižuje. Úpravou této rovnice (jejímž detailům se zde není potřeba věnovat) dostaneme rovnici

$$\Delta w_{ij} = 2 \varepsilon \delta_i F^* a_j,$$

kde δ_i je chyba na uzlu i ve výstupní (tj. rozdíl mezi skutečnou a požadovanou mírou aktivace tohoto uzlu), F^* je derivace aktivační funkce uzlu i podle celkového vstupu $netinput_i$ do uzlu i (tj. sklon aktivační funkce) a a_j je míra aktivace uzlu j ze vstupní vrstvy, ze kterého vede spoj do uzlu i . Vedle míry aktivace a_j , která umožňuje zohlednit míru „spoluviny“ uzlu j za chybu

δ_i při úpravě váženého spoje w_{ij} , je zde klíčová derivace aktivační funkce F^* , neboť z její přítomnosti v rovnici vyplývá, že LMS nelze použít v případech, kdy pro uzly ve výstupní vrstvě perceptronu použijeme binární prahovou funkci, jejíž sklon je buď 0, nebo ∞ . Tyto hodnoty neumožňují vypočítat smysluplnou hodnotu Δw_{ij} . LMS lze naopak velice úspěšně použít pro lineární aktivační funkci, která má spojitý a navíc konstantní sklon. Pro lineární aktivační funkci tak LMS bude mít podobu rovnice

$$\Delta w = k \delta_i a_j,$$

kde k je konstanta, která označuje násobek míry učení ε a derivace lineární aktivační funkce F^* . Procedura vyhledávání extrému (*gradient descent/hill climbing*) tak v případě použití lineární aktivační funkce vede ke stejnému učicímu algoritmu jako je perceptronové konvergenční pravidlo, resp. delta pravidlo ($\Delta w_{ij} = \varepsilon \delta_i a_j$), přestože tato pravidla učení pracují na základě odlišných principů.

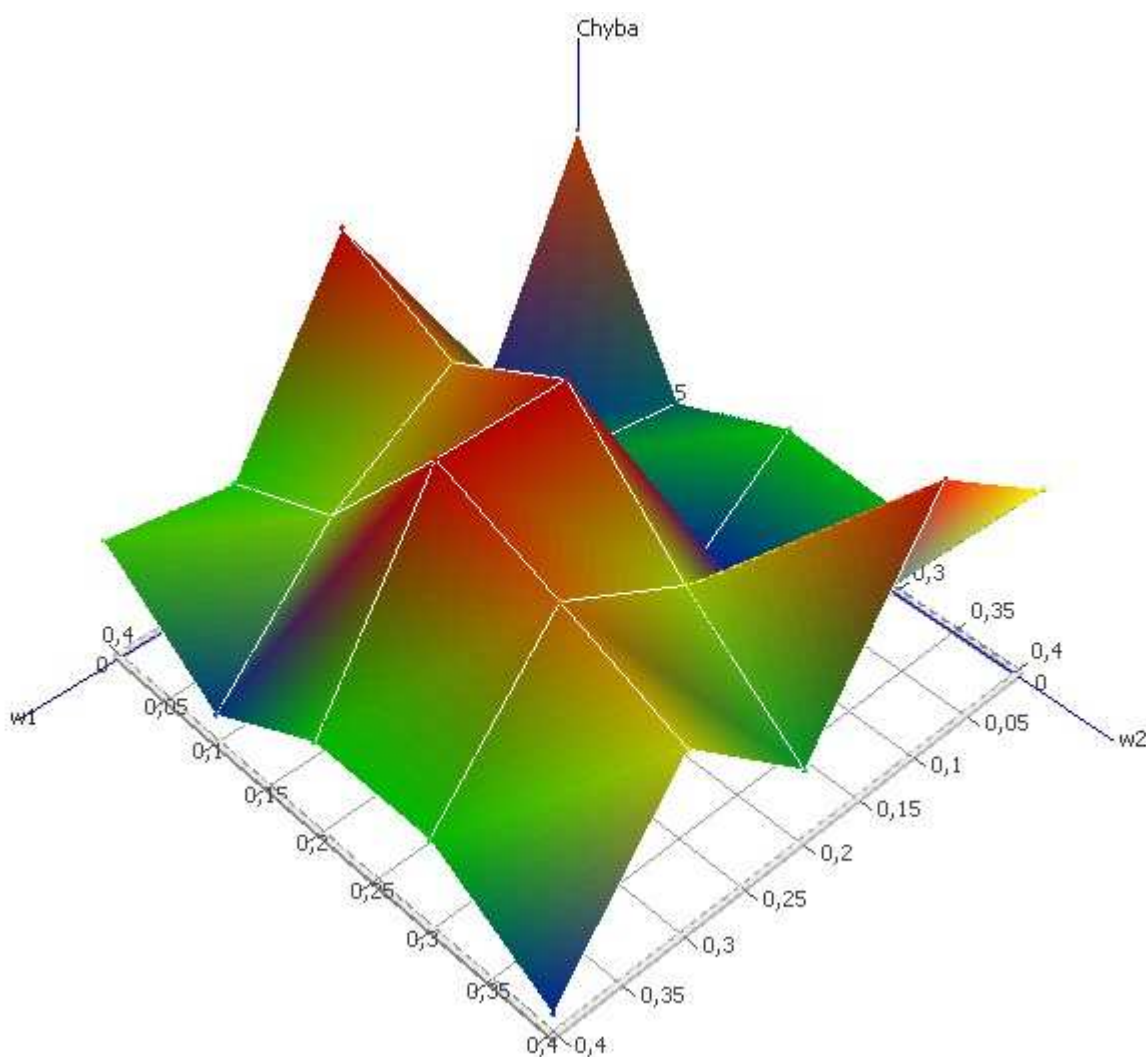
Specifické dopady na proces učení s pomocí LMS má použití biologicky plausibilnější logistické aktivační funkce. Sклон této aktivační funkce má maximální hodnotu v situaci, kdy se celkový vstup ($netinput_i$) do uzlu i pohybuje okolo 0; čím dále se od této hodnoty nacházíme (ve směru plus nebo minus), tím je sklon aktivační funkce menší (viz obrázek 123 na straně 217). To má za následek několik zajímavých a někdy také psychologicky relevantních charakteristik procesu učení umělé neuronové sítě (McLeod, Plunkett, Rolls, 1998):

- Díky tomu, že míra aktivace uzlu s nulovým celkovým vstupem ($netinput_i$) má hodnotu 0,5, procesu učení se mohou „účastnit“ také ty uzly, které na svém vstupu nepřijímají žádné signály od uzlů z předchozí vrstvy neuronové sítě.
- Největší změna vah spojů (tj. největší míra učení) probíhá u uzlů, jejichž celkový vstup se pohybuje okolo 0, tzn. u těch uzlů, které ještě nejsou zapojeny do „pracovního procesu“ transformace vstupů na výstupy. Naopak uzly s velkým celkovým vstupem, kterým je již přiřazena určitá role v „transformačním procesu“, mění váhy svých spojů pouze minimálně.¹
- Učení neuronové sítě probíhá pouze v určitém omezeném časovém období, kdy dochází k postupnému „usazování rolí“ jednotlivých uzlů sítě a po jehož uplynutí je už

¹ Proto se také proces učení umělé neuronové sítě většinou začíná s náhodným nastavením vah v rozmezí $\pm 0,5$, které zaručuje, že průměrný celkový vstup do všech uzlů neuronové sítě se bude pohybovat okolo 0, což zaručuje maximální možnou míru učení.

velice obtížné toto nastavení jakkoli měnit. Tato charakteristika procesu učení neuronové sítě je důležitá z hlediska konekcionistického modelování tzv. *kritických období* v procesu osvojování některých kognitivních funkcí.

Tento učicí algoritmus, jehož aplikaci si lze představit jako pohyb neuronové sítě ve stavovém prostoru definovaném její váhovou maticí a celkovou chybou v její výstupní vrstvě (viz obrázek 132), dokáže (stejně jako perceptronové konvergenční pravidlo, resp. jako delta pravidlo) vždy najít právě to jedno chybové minimum, které charakterizuje chybovou křivku každé jednovrstevné neuronové sítě (disponující pouze jednou vrstvou modifikovatelných vážených spojů). Mohlo by se tak zdát, že máme k dispozici univerzální nástroj, s jehož pomocí můžeme umělou neuronovou síť naučit jakoukoli funkci. Naneštěstí jednovrstevné neuronové sítě nedokáží řešit jednu důležitou skupinu úloh, tzv. *nelineárně oddělitelné*



Obrázek 132: Chybová krajina - stavový prostor umělé neuronové sítě definovaný jejími dvěma váženými spoji a mírou celkové chyby ve výstupní vrstvě. V průběhu procesu učení se neuronová síť pohybuje krajinou plnou kopců a údolí a snaží se zaujmout pozici s co možná nejnížší potenciální energií.

problémy (viz oddíl „2.2.2.3.1 Vícevrstevné neuronové sítě,“). Proto vznikla třída vícevrstevných neuronových sítí, které mají více než jednu vrstvu modifikovatelných vážených spojů, díky čemuž si s tímto typem problémů dokáží snadno poradit. To, s čím by však vícevrstevné neuronové sítě mohly mít problém, jsou tradiční učící algoritmy, které k výpočtu úprav vah spojů používají chybový signál o rozdílu mezi skutečnou a požadovanou mírou aktivace výstupního uzlu i .

2.2.2.2.5 Učící algoritmus zpětného šíření

Podstata problému s učením vícevrstevných neuronových sítí spočívá v tom, že při jejich tréninku máme k dispozici chybovou informaci pouze u uzlů ve výstupní vrstvě, neboť pouze u nich víme, jaká by měla být jejich míra aktivace. V případě uzlů ve skryté vrstvě však žádnou takovou informaci k dispozici nemáme. Tento problém se podařilo vyřešit s pomocí nového učícího algoritmu, tzv. algoritmu zpětného šíření (*backpropagation learning algorithm*), který umožňuje chybové informaci vypočítané ve výstupní vrstvě putovat proti proudu toku aktivačních signálů směrem k uzlům ve skryté vrstvě, které tuto informaci mohou využít k již běžnému výpočtu úprav vah spojů, které je propojují s uzly ve vstupní vrstvě.

Proces šíření chybového signálu směrem od výstupní vrstvy ke skryté vrstvě je tedy zrcadlovým obrazem procesu šíření informace o míře aktivace směrem od vstupní nebo skryté vrstvy k vrstvě výstupní (viz obrázek 133). Dokonce i rovnice popisující výpočet obou typů signálů jsou si podobné. Zatímco míra aktivace uzlu třívrstevné neuronové sítě se počítá podle vzorce

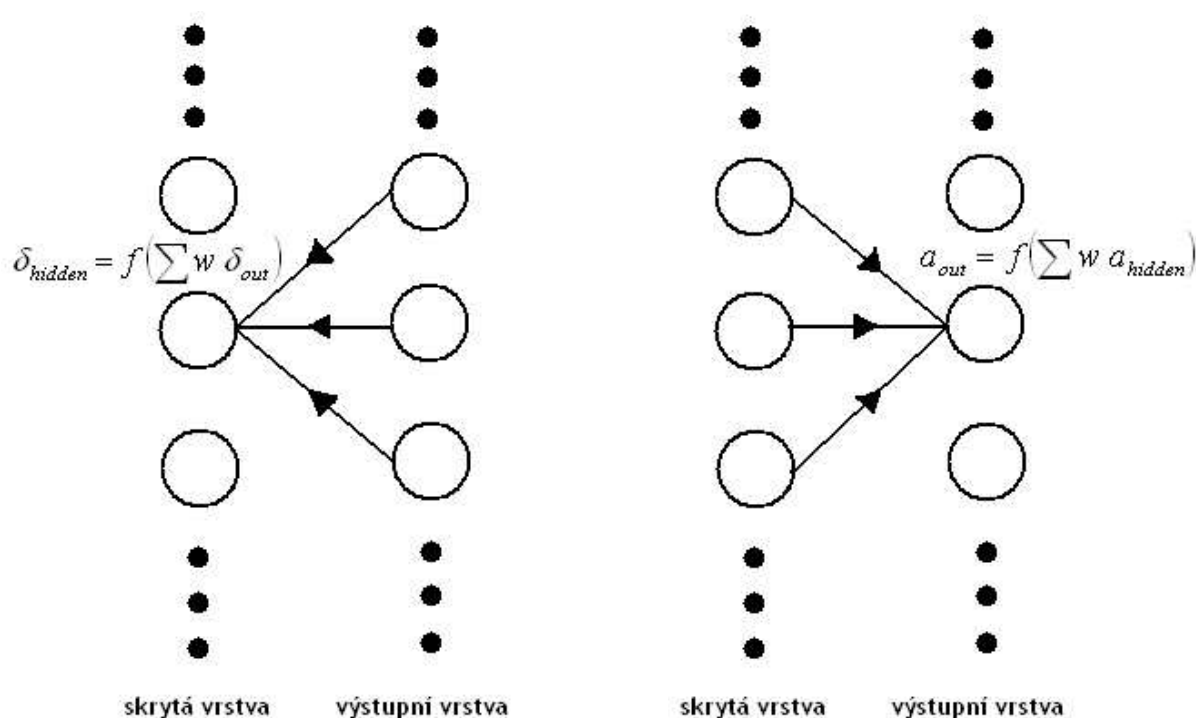
$$a_{out} = f\left(\sum w a_{hidden}\right),$$

kde a_{out} je míra aktivace uzlu ve výstupní vrstvě, f je použitá aktivační funkce, a_{hidden} je míra aktivace uzlů ve skryté vrstvě a w je váha spojů vedoucích z uzlů ve skryté vrstvě do uzlu ve výstupní vrstvě, výpočet chybového signálu pro uzel ve skryté vrstvě má podobu vzorce

$$\delta_{hidden} = f\left(\sum w \delta_{out}\right),$$

kde δ_{hidden} je chyba připisovaná skrytému uzlu, f je určitá chybová funkce, w je váha spojů mezi uzly ze skryté a výstupní vrstvy a δ_{out} je chyba vypočítaná pro jednotlivé výstupní uzly. Jedná se de facto o formální popis procesu připisování „viny“ vybranému skrytému uzlu za chybu

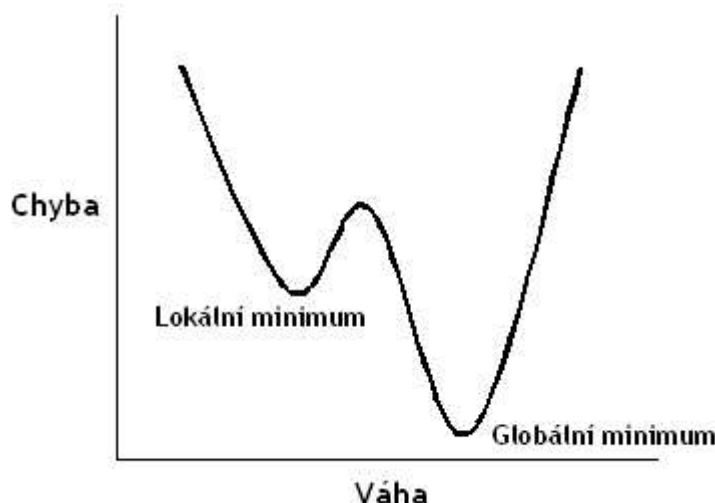
zjištěnou ve výstupní vrstvě. Z tohoto hlediska klíčovou roli hraje údaj o váze spojů mezi uzly ze skryté a výstupní vrstvy: Čím větší je váha spojů, které vybraný skrytý uzel spojují s uzly ve výstupní vrstvě, tím větší je podíl tohoto skrytého uzlu na zjištěné chybě ve výstupní vrstvě, a tím větší také musí být úprava vah spojů vedoucích do skrytého uzlu ze vstupní vrstvy. Úprava těchto spojů se již počítá s pomocí běžného učicího algoritmu, který používá spočítanou chybu pro skrytý uzel δ_{hidden} jako jeden ze svých vstupů. Celý proces zpětného šíření chybového signálu lze iterativně opakovat pro nekonečné množství skrytých vrstev, kdy každá skrytá vrstva přijímá chybový signál od předchozí skryté vrstvy.



Obrázek 133: Srovnání šíření chybového signálu směrem od výstupní vrstvy ke skryté vrstvě (nalevo) a šíření aktivního signálu od skryté vrstvy k vrstvě výstupní. (s úpravami převzato z McLeod, Plunkett, Rolls, 1998, s. 115)

Přestože je teoreticky prokázáno, že vícevrstevná neuronová síť se dokáže naučit jakoukoli funkci, při použití algoritmu zpětného šíření nikdy nemáme jistotu, že se nám podaří najít takovou konfiguraci váhové matice, která by umožnila realizaci požadované funkce. Zatímco tedy perceptronové konvergenční pravidlo zaručuje nalezení řešení pro danou funkci v případě, že takové řešení existuje, algoritmus zpětného šíření nám žádnou takovou záruku nedává, přestože řešení – v podobě specificky nastavené váhové matice – určitě existuje. Důvodem tohoto omezení je tvar chybové křivky vícevrstevných neuronových sítí. Zatímco u jednovrstevných neuronových sítí má taková chybová křivka pravidelný a hladký tvar s

jedním tzv. **globálním minimem** (viz obrázek 131), u vícevrstevných neuronových sítí lze v jejich chybové křivce nalézt několik „proláklin“ a „údolíček“, tzv. **lokálních minim**, která však leží výše než globální minimum (viz obrázek 134). Snadno se tak může stát, že učicí algoritmus využívající strategie vyhledávání extrému uvázne v lokálním minimu, neboť tato strategie nevidí daleko před sebe, ale řídí se pouze lokální zpětnou vazbou v podobě sklonu chybové křivky; a protože na dně lokálního minima je sklon nulový, učicí algoritmus se zde zastaví, neboť se bude „domnívat“, že již dosáhl samého (globálního) dna (minima). Riziko



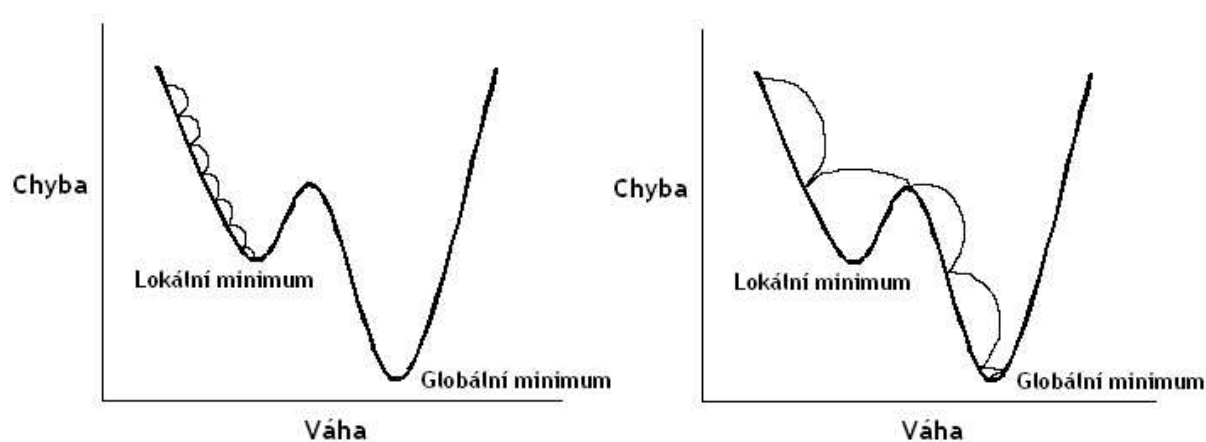
Obrázek 134: Tvar chybové křivky vícevrstevné neuronové sítě. (s úpravami převzato z McLeod, Plunkett, Rolls, 1998, s. 116)

uváznutí v lokálním minimu se zvyšuje při použití nízké míry učení (ϵ), což je parametr, který specifikuje velikost změn vah spojů, a tím tedy také délku „skoků“ po chybové křivce: Čím delší „skoky“, tím větší je šance (nikoli však jistota), že se neuronové sítě podaří lokální minimum „přeskočit“ a že nakonec skončí ve skutečném globálním minimu (viz obrázek 135). Vedle parametru míry učení se k „ošetření“ rizika uváznutí v lokálním minimu používá také tzv. *parametr hybnosti (momentum parameter)* α , který specifikuje, jak velká část úpravy váhy spoje v čase $t-1$ se přičte k úpravě váženého spoje v čase t :

$$\Delta w_{ij(t)} = (\epsilon \delta_i a_j) + \alpha \Delta w_{ij(t-1)}$$

To neuronové sítě umožňuje „prodloužit krok“ a „nabrat rychlost“ i v situacích, kdy nemá možnost se „rozeběhnout z kopce“. Jinak řečeno, vysoký parametr hybnosti umožňuje větší úpravy vah spojů i tehdy, když se neuronová síť nachází blízko dna lokálního minima, kde nízký sklon chybové křivky jinak vede k malým korekcím vážených spojů a ke

zvýšenému riziku uváznutí v lokálním minimu. Vzhledem ke členitosti chybové krajiny vícevrstevných neuronových sítí má na konečný výsledek procesu učení zásadní vliv také konkrétní místo v chybovém prostoru, ze kterého neuronová síť proces učení začíná. Takto se může stát, že zatímco při startu z jedné pozice neuronová síť hledané globální minimum s vysokou pravděpodobností nalezne, při startu z pozice, která je jen trochu odlišná od té první, bude neuronové síti mnohem více hrozit uváznutí v nějakém z mnoha lokálních minim, která se v dané chybové krajině nacházejících. Výchozí bod procesu učení je přitom většinou dán zcela náhodně, neboť také výchozí váhová matice, která definuje výchozí pozici neuronové sítě v chybovém prostoru, je nastavována zcela náhodně (přestože v určitém pevně daném rozmezí).¹



Obrázek 135: Vliv nastavení parametru míry učení (*learning rate*) na riziko uváznutí v lokálním minimu. Nalevo má parametr míry učení nízkou hodnotu, což vede k jen nepatrným úpravám vah spojů a k malým „krokům“ po chybové křivce. Důsledkem je zvýšené riziko uváznutí v lokálním minimu. Při vysoké míře učení (napravo) existuje vyšší pravděpodobnost, že se lokální minimum podaří přeskočit.

2.2.2.2.6 Genetické algoritmy

Z hlediska prevence rizika uváznutí v lokálním minimu se jako dobrá alternativa tradičních učících algoritmů jeví být optimalizační metoda genetických algoritmů. Tato metoda se inspirovuje biologickou evolucí přírodním výběrem, která je de facto druhem prohledávací heuristiky, která v obrovském mnohadimenzionálním genetickém prostoru dokáže najít kvalitní „návod“ na organismy, které jsou schopny úspěšně se vyrovnávat s nároky, které na ně klade vnější prostředí (Dennett, 1995; Hillis, 2003). Evoluce přírodním výběrem, která

¹ Z důvodu reprodukovatelnosti simulačních experimentů většina simulačních programů umožňuje specifikovat tzv. parametr (*seed*), který je součástí algoritmu produkujícího pseudonáhodná čísla, která simulátor používá při generování výchozí váhové matice. Díky tomu lze opakovat sérii simulačních experimentů s různými verzemi neuronové sítě, které mají vždy stejnou výchozí váhovou matici.

vysvětluje vznik nenáhodnosti a uspořádanosti v přírodě, je založena na 1) existenci velké variability organismů, která vzniká především mechanismem náhodných mutací při kopírování genetické informace v průběhu procesu rozmnožování, 2) na tom, že organismy mají schopnost se rozmnožovat a že potomci dědí své vlastnosti po svých rodičích a 3) na tom, že šance organismu na přežití a na úspěšné zplození potomků není otázkou pouhé náhody a štěstí, ale že je z velké části dána určitými vlastnostmi organismu, které mu ve srovnání s jinými organismy umožňují lépe se adaptovat na své životní prostředí.

Přestože je vstup do evolučního algoritmu produktem čistě náhodných mutací, z nichž většina je navíc životu nebezpečná a škodlivá, výstupem samotného evolučního algoritmu jsou nenáhodné, vysoce komplexní a uspořádané organismy dobře přizpůsobené svému prostředí, což je automatický důsledek působení slepého výběrového mechanismu, který jako určitý druh síta propouští dál v čase s mnohem vyšší pravděpodobností ty organismy, jimž náhoda nadělila vlastnosti propůjčující jim z hlediska možnosti přežití a reprodukce v prostředí s omezenými zdroji určitou výhodu nad organismy, k nimž náhoda tak štědrá nebyla.¹ Organismy, které se s požadavky svého životního prostředí vyrovnat nedokáží (nebo to dokáží, ale hůře než jiné organismy), tímto sítem času neprojdou. Celá selekce je tak de facto určitý druh „negativního zpevňování“, které na základě náhodných pokusů a omylů zkouší různé možné varianty životních strategií a všechny neúspěšné pokusy vylučuje tím, že jejich nositelům neumožní mít potomky. Ze zpětného pohledu jsou „všechny [tyto] prostoduché chyby... neviditelné a tak všechno, co vidíme, se jeví jako úchvatná řada triumfů. [...] Tím, že odstraníte z dohledu případy „omylu“, vytvoříte „zázrak“. [Takto] například přes 90 % všech tvorů, kteří kdy žili, zemřelo bez potomků, ale ani jednoho z našich předků nepotkal tento osud. A tak můžeme mluvit o pokolení blažených životů.“ (Dennett, 1996, s. 120-121) Vyjádřeno poněkud expresivněji, všichni jsme „náramní mizerové, i když v poněkud užším technickém smyslu v tom, že [můžeme] vysledovat své předky v dlouhé řadě o něco méně vyvinutých mizerů až k tomu prvnímu samo sebe reprodukujícímu tentononc – které vzhledem k počtu a různorodosti jeho potomstva lze oprávněně popsat jako největšího mizeru všech dob. Každý, kdo nebyl náramný mizera, [je totiž nyní] mrtvý.“ (Stephenson, 2006, s. 15)

¹ Tento paradox, kdy algoritmičtý proces pracující s náhodou vede k nějakému nenáhodnému výsledku, ilustruje Daniel Dennett na aritmetické proceduře dělení velkých čísel: „Vejde se dělenec do dělitele šestkrát, sedmkrát nebo osmkrát? Kdo to ví? Koho to zajímá? Nic z toho nemusíte znát; nemusíte mít ani trochu inteligence a soudnosti na to, abyste zvládli proceduru dlouhého dělení. Algoritmus po vás chce jen to, abyste vybrali nějaké číslo - zcela náhodně, jestli chcete – a zkontrolovali výsledek. Jestliže se vybrané číslo ukáže být příliš malé, zvětšete ho o jedno a začnete znovu; jestliže bude příliš velké, zmenšíte ho. Na proceduře dlouhého dělení je dobrá ta věc, že vás nakonec vždy dovede k cíli a to i tehdy, když jste se při svém první odhadu chovali jako naprostý hlupák, v kterémžto případě vám to bude pouze o něco déle trvat.“ (Dennett, 1995, s. 52-53)

Z hlediska možnosti vzniku vysoce komplexních struktur, které lze nalézt v živé přírodě, hraje klíčovou roli to, že selekce organismů není jednostupňová, ale že má kumulativní povahu: Rozdíl mezi těmito dvěma typy selekce spočívá v tom, že „v *jednostupňové selekci jsou selektované předměty... vybrány jednou provždy. Na druhé straně při kumulativní selekci se tyto předměty „rozmnoužují“ nebo se výsledek jednoho prosívání nějakým způsobem stává zdrojem pro další prosívání, které se opět stává... atd. Předměty jsou selektovány či tříděny v mnoha „generacích“ po sobě. Výsledek jedné generace výběru je výchozím bodem pro další generaci výběru a tak to pokračuje řadu generací.*“ (Dawkins, 2002, s. 56) Kdyby byla evoluce založena na jednostupňové selekci, potom by vznik komplexních struktur živých organismů, například oka, byl stejně astronomicky nepravděpodobný jako náhodné sestavení Boeingu 747 po přeletu hurikánu přes vrakoviště. Bylo by to jako šáhnout do hyperprostorového osudí s miliony dimenzí a z miliard možných návodů, jak dát různé součástky letadla dohromady, hned na první pokus vytáhnout návod na funkční letadlo. Postupná kumulace drobných změn, které jsou dostatečně jednoduché na to, aby mohly vzniknout náhodou, však umožňuje v rozlehlém a mnohazměrném prostoru - definovaném všemi možnými způsoby uspořádání genetické informace - hledat, nalézat a systematicky sledovat těch několik málo vývojových trajektorií, které se skládají z životaschopných organismů.¹ Klíčová je zde podmínka pouze drobných a nepatrných změn, neboť v případě velkých mutačních skoků „*je počet možných míst dopadu astronomicky vysoký [a protože] počet způsobů, jak být mrtvý, je mnohem větší než počet způsobů, jak být živý, má velký náhodný mutační skok v genetickém prostoru velkou šanci, že skončí smrtí. Dokonce i malý náhodný skok má docela slušnou šanci na takový konec. Čím je však skok menší, tím větší má naději na úspěch a na to, že povede k vylepšení.*“ (Dawkins, 2002, s. 84) Při dostatečném množství času tak může vzniknout i něco tak hodně nepravděpodobného jako je lidské oko nebo lidský mozek.

Z hlediska možnosti využití evolučního algoritmu k prohledávání váhové prostoru neuronové sítě je důležité, že evoluce přírodním výběrem je nezávislá na biologickém substrátu, tzn. že ho lze aplikovat i na jiné objekty než jsou organické molekuly DNA. Proces evoluce takto může probíhat v jakémkoli prostředí, které splňuje tři velice obecně specifikované podmínky, kterými jsou: 1) variabilita selektovaných entit, 2) rozmnožování

¹ „Skupina živočichů, kteří kdy na Zemi žili, je jen nepatrnou podskupinou všech živočichů, kteří teoreticky mohli existovat. Tito skutečně existující živočichové jsou výsledkem velmi malého počtu evolučních trajektorií, jež vedou genetickým prostorem. Drtivá většina těchto cest by dala vzniknout nemožným příšerám. Skuteční živočichové jsou roztroušeni tu a tam mezi hypotetickými monstry a každý z nich si sedí na svém jedinečném místě v hyperprostoru. Každý skutečný tvor je obklopen malým shlukem sousedů, z nichž většina nikdy neexistovala. Pár z nich jsou ale jeho předkové, potomci a bratraci.“ (Dawkins, 2002, s. 84)

těchto entit a dědičnost jejich vlastností a 3) rozdílné přežívání entit v závislosti na jejich zdatnosti (*fitness*), která je dána interakcí jejich vlastností s charakteristikami životního prostředí (Dennett, 1995, s. 343). Dawkins (1998, s. 174) stejně obecně charakterizuje základní princip evoluce jako rozdílné přežívání replikátorů. Těmi přitom mohou být jakékoli entity, které mají schopnost se rozmnožovat. Známým příkladem nebiologického replikátoru jsou memy - Dawkinsem (1998) postulované základní jednotky kulturní evoluce, které se v prostředí tvořeném lidskými mozky chovají podobně jako geny v prostředí biologickém, tzn. že se snaží maximalizovat své šance na přežití šířením co možná největšího počtu svých kopií.

Již od 50. let 20. století se objevovaly snahy využít principů evoluce k optimalizaci řešení řady různých technických problémů. V 60. letech 20. století *John Holland* vytvořil teoreticky propracovaný výpočetní model evoluce - tzv. genetické algoritmy (dále GA), které umožňují úspěšně modelovat evoluci umělých *chromozomů* tvořených řetězcí bitů, které mohou reprezentovat různé umělé organismy nebo alternativní návrhy řešení nějakého problému, jakým je například optimalizační úloha nastavení váhové matice umělé neuronové sítě tak, aby chyba na její výstupní vrstvě byla co nejmenší. Každý chromozom se skládá z několika *genů*, respektive z několika *alel* – konkrétních verzí genu kódujícího jednu z dílčích charakteristik umělého organismu, které determinují jeho relativní zdatnost ve vztahu k selekčnímu tlaku vyvíjenému na něj okolním prostředím (včetně ostatních umělých organismů). V případě umělé neuronové sítě by taková alela reprezentovala konkrétní váhu jednoho spoje mezi dvěma uzly. Celý chromozom by pak reprezentoval celou váhovou matici. Podstata GA spočívá v aplikaci několika tzv. *genetických operátorů* na populaci většího množství chromozomů v diskrétních časových krocích, které se iterativně opakují mnohokrát za sebou. Tyto diskrétní časové kroky evolučního procesu se nazývají „*generace*“. Výchozí populace, tj. první generace chromozomů je generována čistě náhodně, každá další pak vzniká aplikací genetických operátorů na předchozí generaci chromozomů. Součástí základního souboru genetických operátorů jsou operátory *selekce*, *crossover*, *mutace* a někdy také *inverze*. Při aplikaci operátoru selekce se z dané populace chromozomů vybírají ty chromozomy, kterým bude umožněna reprodukce, tj. kterým bude umožněno, aby předaly část svých genů do další generace chromozomů. Tento výběr se děje na základě výpočtu zdatnosti jednotlivých chromozomů. Ta je dána interakcí vlastností kódovaných jednotlivými geny chromozomu s nějakým uměle nastaveným selekčním kritériem, kterým je v případě trénování či učení umělé neuronové sítě chyba na výstupní vrstvě: Čím menší chyba, tím větší zdatnost dané váhové matice a tím větší pravděpodobnost, že se chromozomu kódujícímu tuto

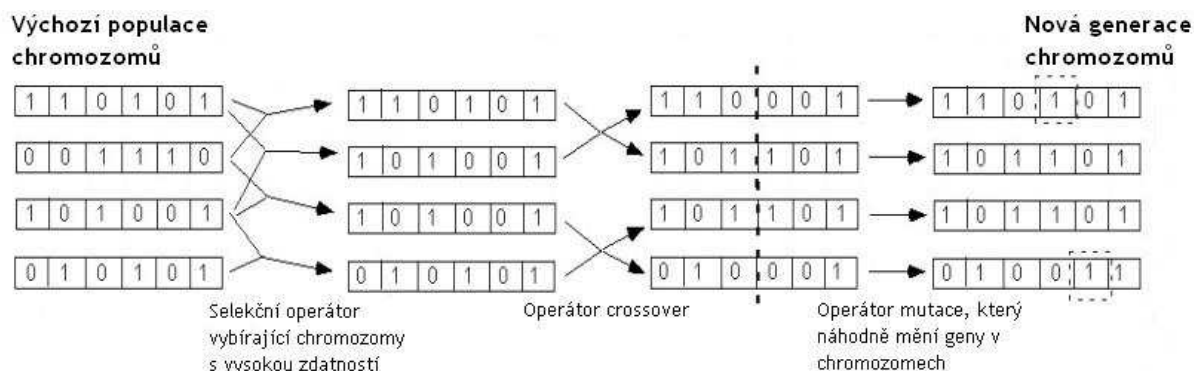
váhou matici podaří předat část svých genů další generaci chromozomů. Operátor selekce lze implementovat různými způsoby. Například je možné eliminovat 50 % populace chromozomů s nejnižší zdatností a reprodukci umožnit zbývajícím 50 % populace; jiný způsob by spočíval v reprodukci všech chromozomů s tím, že počet jejich potomků by byl přímo úměrný míře jejich zdatnosti atp. Operátor crossover kříží geny z každých dvou vybraných chromozomů a vytváří tak novou generaci chromozomů, které mají podobné vlastnosti jako jejich rodičovské chromozomy. Crossover, který napodobuje rekombinaci genů, ke které dochází při sexuálním rozmnožování, probíhá tak, že se na dvou chromozomech náhodně zvolí bod (*crossover point*), podél kterého se oba chromozomy rozdělí na dvě části, z nichž si jednu chromozomy mezi sebou vymění. Operátor mutace náhodně mění hodnotu některého z míst na chromozomu; dnes již příliš nepoužívaný genetický operátor inverze pak prohazuje pořadí dvou sousedících genů. Výsledkem aplikace všech těchto genetických operátorů je nová generace chromozomů¹, které slouží jako vstup do dalšího kola evolučního algoritmu (Franklin, 2001; Fanta, 2001; Mitchell, Forrest, 1994). Celý GA tak ve své nejjednodušší podobě (bez genetického operátoru inverze) může mít podobu čtyř následujících kroků (viz také obrázek 136):

1. Začni s náhodně vygenerovanou populací chromozomů.
2. Pro každý chromozom v populaci spočítej jeho zdatnost.
3. Vytvoř novou generaci chromozomů aplikací selekčního operátoru a operátorů crossover a mutace na současnou generaci chromozomů.
4. Jestliže bylo dosaženo stanoveného kritéria, ukonči činnost, jinak přejdi zpět ke kroku č. 2.

Prostřednictvím tohoto jednoduchého algoritmu lze „vyšlechtit“² požadovanou váhovou matici, která je „jako šitá na míru“ funkci, kterou chceme neuronovou síť naučit. Díky tomu, že GA na rozdíl od klasických učících algoritmů vždy pracuje s populací většího množství chromozomů kódujících různé váhové matice, GA dokáže najednou prozkoumávat větší část váhového prostoru neuronové sítě a hlavně má možnost najednou porovnávat větší množství

¹ Počet chromozomů v populaci v průběhu jednotlivých generací většinou nekolísá a je udržován na určité konstantní, předem stanovené úrovni.

² Při formulaci své teorie evoluce přírodním výběrem se šlechtitelstvím inspiroval také Charles Darwin, který přírodní výběr chápal jako analogii výběru, který provádí chovatel za účelem vyšlechtění organismu s požadovanými vlastnostmi. Rozdíl mezi přírodním výběrem a chovatelem je pouze v tom, že přírodní výběr je „slepý“, tzn. že při selekci organismů před sebou nemá žádný vzdálený cíl, kterého by chtěl dosáhnout. Jednoduše vybírá pouze ty organismy, které se dokáží úspěšně adaptovat na podmínky momentálně panující v životním prostředí. Přírodní výběr je zde přitom pouze metafora (převzatá právě z chovatelství) pro skutečnost, že některé organismy kvůli svým vlastnostem zahynou dříve než jiné a že kvůli tomu po sobě nezanechají žádné potomstvo.



Obrázek 136: Jednotlivé kroky genetického algoritmu. (převzato z nápovědy k programu *OptWorks Excel*)

možných lokálních minim, čímž se zvyšuje pravděpodobnost, že se neuronová síť vyhne lokálnímu minimu a že narazí na skutečné globální minimum. Riziko uváznutí v lokálním minimu snižuje také prvek náhody v GA v podobě genetického operátoru mutace, který neuronové síti umožňuje v jejím váhovém prostoru objevovat nová a dosud netušená „území“. Chyby a mutace jsou takto stejně jako v případě biologické evoluce „prameny novosti [a] jedinou příležitostí se něčemu [skutečně novému] naučit“ (Dennett, 1996, s. 117-118).¹

Nevýhodou GA je jejich poměrně vysoká výpočetní náročnost. Jako učící algoritmus se proto GA dají použít pouze u jednodušších neuronových sítí s menším počtem uzlů a s menším počtem vážených spojů mezi nimi.² Z hlediska modelování kognitivních funkcí je nevýhodou GA také jejich malá biologická plausibilita. Přestože se objevilo několik teorií neuronového darwinismu či darwinistických modelů lidské mysli (Edelman, 1987; Calvin 2000), mezi kognitivními vědci v současné době převládá názor, že darwinovská selekce není primárním mechanismem stojícím v základech lidské kognice.

Biologickou neadekvátnost lze vytknout také učícímu algoritmu zpětného šíření, který předpokládá možnost šíření chybového signálu proti směru šíření signálu aktivačního, přestože v mozku axony umožňují šíření signálu pouze jedním směrem. Vystává tak otázka,

¹ Prostřednictvím postupného snižování pravděpodobnosti mutace je možné napodobit také jiný optimalizační algoritmus, tzv. **simulované žhání** (*simulated annealing*). To se inspirované metalurgickou technikou žhání, které se používá k docílení optimálních vlastností kovů a které spočívá v rychlém zahřívání kovu a jeho pomalém ochlazení, během kterého mají atomy – vysokou teplotou vytržené ze svých původních pozic v krystalické mřížce – možnost usadit se v energeticky výhodnější konfiguraci znamenající lepší vlastnosti kovového materiálu. V případě umělé neuronové sítě a GA to znamená, že na počátku vysoká pravděpodobnost mutace neuronové sítě umožňuje prozkoumat co možná nejvíce různých oblastí váhového prostoru; s postupem času, jak se zvyšuje šance, že se již síť nachází někde poblíž globálního minima, se míra mutací snižuje, aby se váhová matice mohla v tomto globálním minimu úspěšně usadit.

² Na příloženém CD je možné nalézt excelovský soubor *UNS_GA_XOR*, který představuje ukázkou aplikace optimalizační metody genetických algoritmů v rámci nastavení váhové matice jednoduché umělé neuronové sítě implementující logickou funkci XOR. K vlastním pokusům s GA je potřeba si nejdříve nainstalovat 30-denní trial verzi doplňku k Microsoftu Excel v podobě *OptWorks Excel*, který se na příloženém CD nachází pod názvem *OptWorks_PC_Demo*.

do jaké míry lze konekcionistické modely trénované pomocí algoritmu zpětného šíření považovat za skutečné modely vybraných kognitivních funkcí. Kognitivní vědci jsou v tomto ohledu rozděleni do dvou skupin: Jedna skupina algoritmus zpětného šíření obhajuje s poukazem na to, že důležité je to, co se neuronová síť naučí, tedy její reprezentace, a nikoli konkrétní způsob, jakým si je osvojila (ve smyslu konkrétní podoby obecnějšího učícího algoritmu pracujícího na principu vyhledávání extrému (gradient descent/hill climbing)). Ve prospěch algoritmu zpětného šíření hovoří také skutečnost, že mnohé konekcionistické modely trénované pomocí tohoto algoritmu vykazují vzorec osvojování nebo naopak ztráty dané kognitivní funkce, který se velice podobá tomu pozorovanému u člověka. Druhá skupina kognitivních vědců používání algoritmu zpětného šíření naopak odmítá a při modelování kognitivních funkcí šáhá spíše po biologicky plausibilnějším Hebbově učícím algoritmu, u něhož jsou již celkem dobře známy biologické mechanismy, na základě kterých by mohl fungovat i v lidském mozku. V některých případech může být také diskutabilní psychologická plausibilita všech učících algoritmů pracujících s chybovým signálem počítaným pro uzly ve výstupní vrstvě; tato pochybnost je oprávněná zejména v případě těch kognitivních funkcí, u kterých není prokázáno, že by v průběhu jejich osvojování člověk, resp. jeho mozek dostával nějakou formou zpětnou vazbu o svém aktuálním výkonu, tak aby ji mohl využít k úpravě synaptických spojení mezi neurony.

2.2.2.3 Vliv architektury neuronové sítě na způsob zpracovávání informací

Vedle váhové matice umělé neuronové sítě má na způsob zpracovávání informací zásadní vliv také její architektura, tedy počet uzlů a zejména pak způsob jejich zapojení.

2.2.2.3.1 Vícevrstevné neuronové sítě

Jak již bylo uvedeno v předchozím oddíle, neuronové sítě pouze s jednou vrstvou modifikovatelných vážených spojení (tzv. *jednovrstevné neuronové sítě* či *perceptrony*) nedokáží řešit jednu důležitou skupinu úloh, tzv. **nelineárně oddělitelné problémy** (*nonlinearly separable problems*). Tato slabina je pouze odvrácenou stranou jejich jinak silné stránky, která spočívá v obecné schopnosti umělých neuronových sítí generalizovat své poznatky zakódované ve váhové matici i na neznámé případy, které nebyly součástí jejich tréninkového prostředí. Tato jejich schopnost je vedlejším efektem distribuovaného způsobu reprezentace informací, kdy i jen částečné překrytí dvou různých aktivačních vzorců

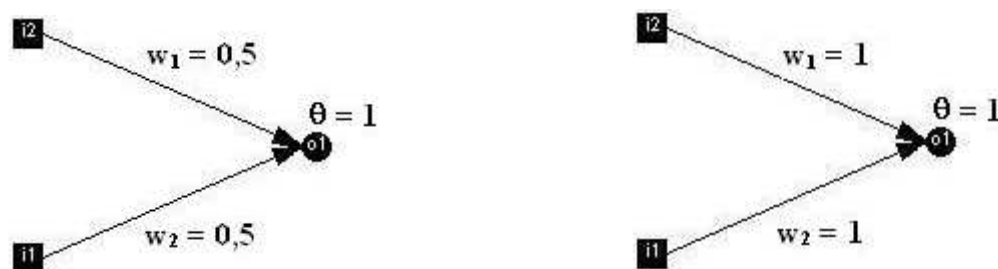
kódujících dva různé podněty (například [1 0 1 0 1 0 0 1 0] a [0 0 1 0 1 0 0 1 0]) umožňuje umělé neuronové síti reagovat na oba podobným způsobem. Jedná se de facto o formu analogického myšlení, které vychází z obecného pravidla, že „u podobných příčin čekáme podobné účinky“ (Hume, 1972, s. 68). Toto pravidlo je v mnoha případech užitečným návodem, jak adaptivně reagovat v nových a neznámých situacích, se kterými nemáme žádnou osobní zkušenost. Takto může například makléř vnímat současnou situaci na finančních trzích jako podobnou té před pěti lety a per analogiam tak dojít k závěru, že stejně jako před pěti lety by mohlo být výhodné investovat finanční prostředky do určitého specifického investičního nástroje, například do derivátů. Naneštěstí ale existuje neméně případů, kdy je toto pravidlo zavádějící a kdy spoléhání se pouze na povrchní podobnost mezi dvěma podněty vede k nesprávným závěrům a k nesprávným rozhodnutím. Například to, že žampion a muchomůrka bílá mají velice podobný vzhled ještě neznamená, že by muchomůrka bílá byla také stejně jedlá jako žampion. V tomto případě se jako užitečná a adaptivní jeví být naopak schopnost pojímat a chápat dva podobné podněty jako rozdílné.

Omezení jednovrstevných neuronových sítí spočívá v tom, že jsou schopny generalizace pouze na základě této povrchní podobnosti mezi vstupními vzorci aktivace. Jinými slovy, jednovrstevné neuronové sítě jsou schopny řešit pouze lineárně oddělitelné problémy. Tato jejich „slepá skvrna“ se často ilustruje na booleovské funkci „xor“ (*exclusive or*), která na rozdíl od booleovských funkcí „a“ a „nebo“ představuje nelineárně oddělitelný problém. Booleovské funkce jsou de facto určitým druhem klasifikační úlohy, která spočívá v rozřazování čtyř dvoubitových vstupů (00, 10, 01, 11) do dvou různých kategorií (1, 0). V tabulce 8 jsou uvedeny všechny čtyři možné vstupy a tři klasifikační pravidla pro jejich přiřazování do dvou kategorií v závislosti na tom, zda se jedná o funkci „a“, „nebo“ či „xor“.

Vstup	Výstup		
	"a"	"nebo"	"xor"
0 0	0	0	0
1 0	0	1	1
0 1	0	1	1
1 1	1	1	0

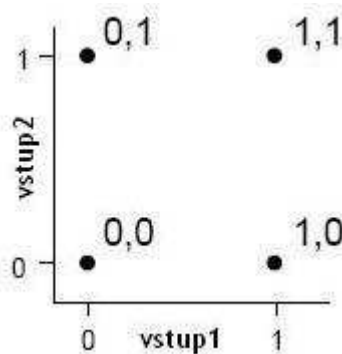
Tabulka 8: Booleovské funkce „a“, „nebo“ a „xor“.

Booleovské funkce „a“ a „nebo“ lze velice snadno realizovat pomocí jednovrstvé neuronové sítě s binární prahovou aktivační funkcí, která se skládá ze dvou uzlů ve vstupní vrstvě a z jednoho uzlu ve vrstvě výstupní. Dva konkrétní příklady takové neuronové sítě se specificky nastavenými váhami spojů a specificky nastavenou prahovou hodnotou výstupního uzlu jsou zachyceny na obrázku 137. Zatímco pro funkce „a“ a „nebo“ je snadné v rámci jednovrstvé neuronové sítě nalézt potřebné nastavení váhové matice a prahové hodnoty, pro booleovskou funkci „xor“ žádnou takovou kombinaci váhové matice a prahové hodnoty



Obrázek 137: Dvě jednovrstvé neuronové sítě s binární prahovou aktivační funkcí implementující booleovskou funkci „a“ (nalevo) a „nebo“ (napravo). (vytvořeno s pomocí simulačního softwaru Tlearn)

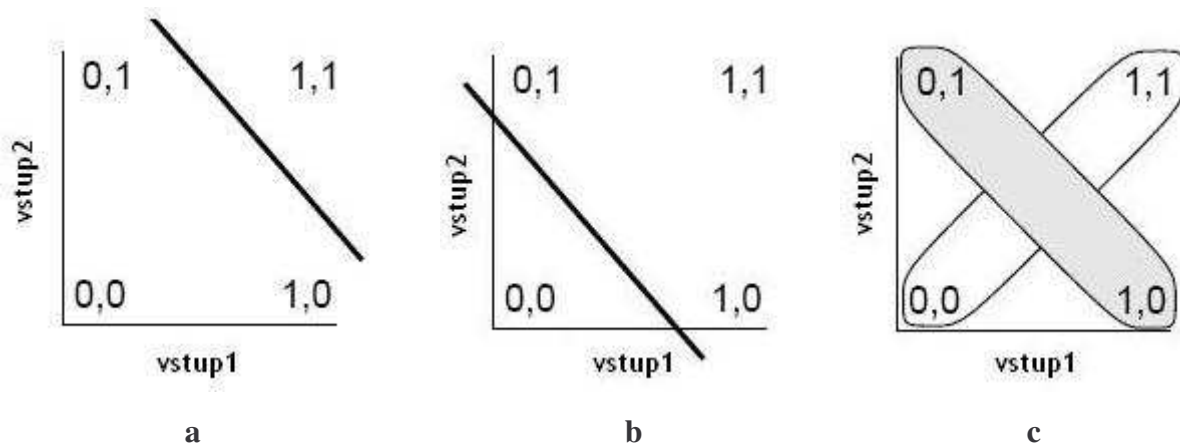
nalézt nelze. Důvodem je struktura podobnosti vstupního reprezentačního prostoru definovaného vstupními aktivačními vektory. Na obrázku 138 je tento reprezentační prostor zobrazen v podobě jednoduchého 2-D grafu, kde osy x a y odpovídají míře aktivace dvou vstupních uzlů. Jednotlivé vstupy jsou zde znázorněny jako body, jejichž pozice je dána



Obrázek 138: Vstupní reprezentační prostor booleovských funkcí. (s úpravami převzato z Elman a kol., 1996, s. 62)

kombinací hodnot na osách x a y . Klasifikaci vstupních aktivačních vzorců v závislosti na použité booleovské funkci pak lze v grafu znázornit pomocí přímky, která vstupy rozděluje do dvou kategorií (1/0). Na obrázku 139a a 139b je takto znázorněno rozdělení reprezentačního prostoru pro booleovskou funkci „a“ a „nebo“. Přesná pozice přímky je zde dána specifickým

nastavením váhové matice neuronové sítě a prahové hodnoty výstupního uzlu. V případě booleovských funkcí „a“ a „nebo“ existuje mnoho způsobů, jak přímkou prostorem proložit tak, aby všechny čtyři vstupy byly odpovídajícím způsobem klasifikovány do dvou kategorií.



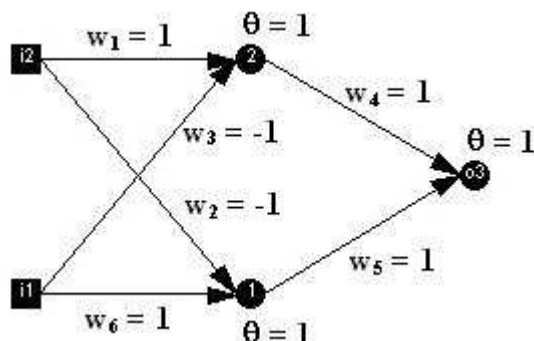
Obrázek 139: Grafické znázornění booleovských funkcí a) „a“, b) „nebo“ a c) „xor“. (s úpravami převzato z Elman a kol., 1996, s. 63)

Právě tato možnost rozdělit reprezentační prostor pomocí jedné přímky¹ na dvě části činí z funkcí „a“ a „nebo“ lineárně oddělitelný problém. To, co funguje pro funkce „a“ a „nebo“ však nefunguje pro booleovskou funkci „xor“: Neexistuje žádný způsob, jak přímkou proložit reprezentačním prostorem tak, aby se vstupy 00 a 11, resp. 01 a 10 nacházely ve stejné kategorii (viz obrázek 139c). Z hlediska nastavení váhové matice příčina spočívá v tom, že oba vážené spoje musí být pozitivní a dostatečně silné na to, aby při prezentaci vstupů 10 a 01 neuronová síť na výstupu vyprodukovala 1. To však nevyhnutelně vede k tomu, že 1 se na výstupu neuronové sítě objeví i v případě vstupu 11, který ve skutečnosti patří - spolu se vstupem 00 - do druhé kategorie (0). Podstata problému tedy tkví v tom, že vstup 11 se neuronové síti jeví jako podobnější (bližší) vstupům 10 a 01 než vstupu 00 (viz také obrázek 138), takže neuronová síť pak na vstup 11 reaguje stejně jako na vstupy 10 a 01 a nikoli jako na vstup 00.

Řešení spočívá v přidání jedné vrstvy uzlů navíc (viz obrázek 140). Tato tzv. *skrytá vrstva* umožňuje neuronové síti si vytvářet **vnitřní reprezentace** podnětů, které se mohou zásadním způsobem lišit od způsobu, jakým jsou tyto podněty reprezentovány ve vstupní vrstvě. Vícevrstevná neuronová síť takto může prostřednictvím svých skrytých uzlů vnitřně re-representovat vstupní podněty a změnit strukturu podobnosti reprezentačního prostoru, ze kterého neuronová síť vychází při transformaci vstupů na výstupy. V případě funkce „xor“

¹ V případě 3-D reprezentačního prostoru by dělící hranici mezi dvěma kategoriemi tvořila rovina; v případě čtyř a vícedimenzionálního reprezentačního prostoru by tuto roli hrála hyperrovina.

jsou vstupní aktivační vzorce přereprezentovány ve skryté vrstvě tak, aby vstup 11 byl podobnější (bližší) vstupu 00 než vstupům 10 a 01 (viz tabulka 9). V této podobě se již problém stává řešitelným i pro jednovrstevnou neuronovou síť, zde tvořenou spoji a uzly



Obrázek 140: Třívrstevná neuronová síť úspěšně implementující booleovskou funkci „xor“. (vytvořeno s pomocí simulačního softwaru Tlearn)

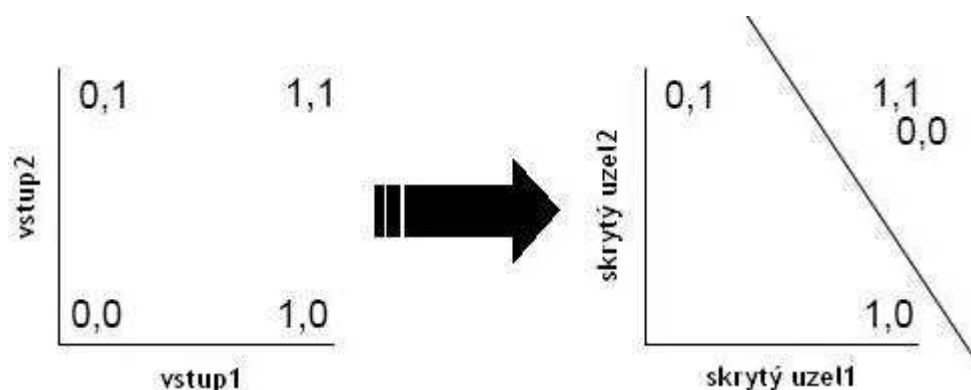
skryté a výstupní vrstvy. Skrytá vrstva takto z nelineárně oddělitelného problému dělá problém lineárně oddělitelný (viz obrázek 141). Skrytá vrstva de facto funguje jako jistý druh „gravitační čočky“, která deformuje vnitřní obraz z vnějšku působících podnětů způsobem, který je užitečný a žádoucí z hlediska řešení daného problému (viz obrázek 140). Skrytá vrstva takto neuronovou síť osvobozuje od diktátu její „vrozené“ tendence reagovat na

Vstup	Skrytá vrstva		Výstup
	skrytý uzel1	skrytý uzel2	
0 0	0	0	0
1 0	1	0	1
0 1	0	1	1
1 1	0	0	0

Tabulka 9: Aktivace uzlů ve vstupní, skryté a výstupní vrstvě neuronové sítě implementující funkci „xor“.

podněty na základě jejich povrchní podobnosti. To neznamená, že by se tímto svého „zlozvyku“ zbavila nadobro, pouze jí to v případě potřeby dovoluje její „nutkavou potřebu“ neutralizovat a své reakce a soudy založit na abstraktnějších vztazích mezi jednotlivými vstupními aktivačními vzorci než je jejich „fyzická“ podobnost.

Z hlediska schopnosti vícevrstevné neuronové sítě generalizovat své poznatky na nové a dosud nepotkané případy hraje klíčovou roli počet uzlů ve skryté vrstvě. V případě, že je ve skryté vrstvě příliš mnoho uzlů, dochází často k tzv. **přeučení** (overlearning či overfitting), kdy se neuronová síť na základě prezentovaných příkladů neučí nějakému obecnějšímu pravidlu, které by bylo možné aplikovat i na nové případy, ale pouze se prezentované příklady učí mechanicky nazpaměť. Důvodem toho je to, že s příliš mnoha uzly ve skryté vrstvě



Obrázek 141: Grafické znázornění deformace struktury podobnosti reprezentačního prostoru. Nalevo je zachycen způsob reprezentace vstupů ve vstupní vrstvě a z něj vyplývající vzdálenosti (podobnosti) mezi jednotlivými vstupy. Napravo je pak zobrazen způsob reprezentace vstupů ve skryté vrstvě. Oproti vstupní vrstvě má vstup 11 ve skryté vrstvě mnohem blíže ke vstupu 00, se kterým patří do stejné kategorie, než ke vstupům 10 a 01, které patří do druhé společné kategorie. Díky této deformaci struktury podobnosti reprezentačního prostoru se z booleovské funkce „xor“ stává lineárně oddělitelný problém, který lze vyřešit i s pomocí jednovrstevné neuronové sítě. (s úpravami převzato z Elman a kol., 1996, s. 65)

neuronová síť dostává k dispozici rozlehlý mnohadimenzionální reprezentační prostor, ve kterém lze snadno umístit každý z prezentovaných podnětů jako samostatnou kategorii, která s ostatními nemá nic společného. Výsledkem je pak to, že každý z prezentovaných podnětů je neuronovou sítí vnímán jako „idiograficky“ jedinečný a neopakovatelný jev, jehož vlastnosti a charakteristiky nelze vztahovat k ničemu jinému než k němu samotnému. Fanta (2000, s. 24-25) tento stav neuronové sítě charakterizuje jako situaci, kdy váhová matice „neodráží skrytou funkci dat optimálně: Množina dat užitá k trénování neuronové sítě vyjadřuje vztahy mezi vstupy příliš dobře, takže neponechává prostor pro jiná data obsahující šumy.“

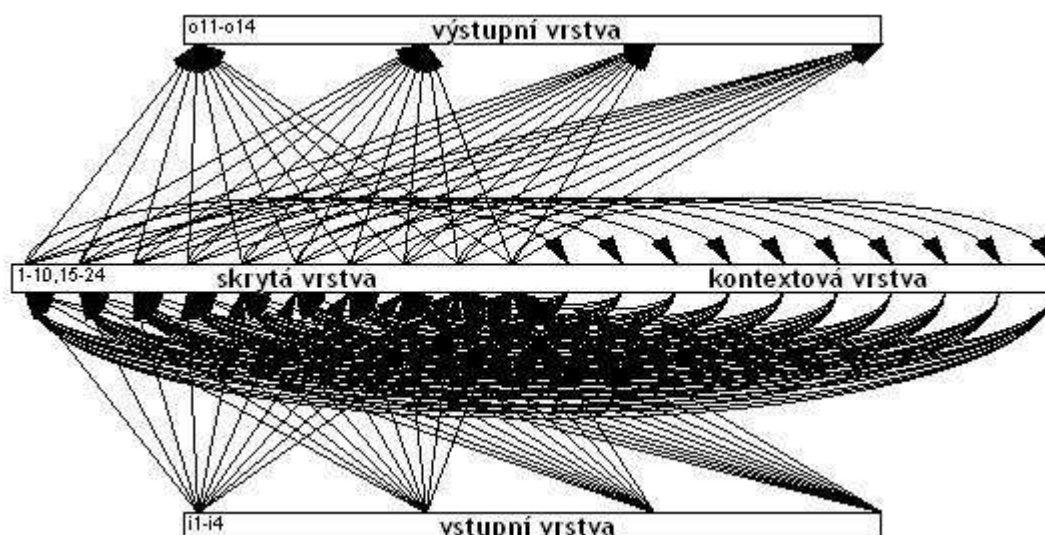
Řešením je snížení počtu uzlů ve skryté vrstvě, kdy se velikost reprezentačního prostoru zmenší. Aby neuronová síť byla i nadále schopna realizovat požadovanou transformaci vstupů na výstupy, musí najít efektivnější způsob reprezentace podnětů za využití nějaké „zkratky“ v podobě hlubších a abstraktnějších vztahů existujících mezi jednotlivými vstupními podněty - jedná se de facto o formu komprese dat, která využívá v datech existujících zákonitostí.

Tímto způsobem je neuronová síť donucena si osvojit nějaké obecnější pravidlo či zákonitost, kterou pak může využít i při zpracovávání nových a neznámých podnětů. Při redukci počtu uzlů ve skryté vrstvě však hrozí, že se překročí určitá kritická hranice, za kterou již neuronová síť nemá k dispozici dostatečnou paměťovou a výpočetní kapacitu (která u umělých neuronových sítí v podobě vah spojů „jedno je“), aby si dokázala osvojit a zapamatovat pravidlo implicitně obsažené v prezentovaných příkladech (tzv. *underlearning* či *underfitting*). Při volbě počtu uzlů ve skryté vrstvě je tak tedy potřeba zvolit zlatou střední cestu, kdy uzlů není ani příliš mnoho, ani příliš málo, ale právě tak akorát.

2.2.2.3.2 Jednoduché rekurentní síť

Možnosti klasické dopředné neuronové sítě, ve které se signály mohou šířit pouze jedním směrem (a to od vstupní vrstvy před vrstvou skrytou k vrstvě výstupní), značně rozšiřuje doplnění její konektivity o různé zpětnovazební (rekurentní) spoje mezi uzly z jednotlivých vrstev. Příkladem neuronové sítě s tímto typem konektivity je jednoduchá rekurentní síť (*simple recurrent network*, dále SRN), jejíž součástí je vedle běžné vstupní, skryté a výstupní vrstvy rovněž tzv. kontextová vrstva (viz obrázek 142). Uzly ze skryté vrstvy jsou spojeny s uzly z kontextové vrstvy souborem nemodifikovatelných spojů, jejichž váha má pevně danou hodnotu, tak aby kontextová vrstva mohla plnit úlohu paměti, která uchovává informaci o aktuální míře aktivace uzlů ve skryté vrstvě. Tuto informaci uzly v kontextové vrstvě v následujícím časovém kroku odesílají prostřednictvím souboru (již modifikovatelných) spojů zpět do skryté vrstvy. Uzly ve skryté vrstvě tak zároveň zpracovávají informace přicházející k nim ze vstupní vrstvy a informace, které jsou výsledkem jejich vlastní aktivity z předchozího časového kroku, kdy zpracovávaly tehdy aktuální informaci ze vstupní vrstvy. Díky tomu, že se celý cyklus opakuje neustále dokola, odráží se v aktuální aktivitě uzlů ve skryté vrstvě informace ze všech předešlých časových kroků. Například v čase $t + 1$ budou uzly ve skryté vrstvě zpracovávat informace jak z času $t + 1$, tak i z času t ; v čase $t + 2$ to budou informace z času $t + 2$, $t + 1$ a t atd. Tento způsob zpracovávání informací umožňuje SRN implementovat funkce, které nezávisí pouze na aktuálním vstupu, ale také na dřívější historii vstupů.¹ Přidání kontextové vrstvy je tak jakýmsi ekvivalentem neurochirurgického zákroku, který pacienta zbavuje anterográdní

¹ SRN je tak jakousi obdobou konečného automatu (viz strana 54-55). Byl také podán formální důkaz, že rekurentní síť tohoto typu jsou skutečně ekvivalentní Turingovu stroji, a že tak tedy disponují stejnou univerzální výpočetní silou jako UTS (Siegelmann, Sontag, 1995).



Obrázek 142: Jednoduchá rekurentní síť. (vytvořeno s pomocí simulačního softwaru Tlearn)

amnésie a nutnosti žít pouze *hic et nunc*, tedy zde a nyní. Praktickým důsledkem je pak to, že jestliže v posloupnosti tréninkových příkladů existuje nějaká statistická zákonitost, SRN ji dokáže najít a použít v rámci nějaké anticipační úlohy, kde je potřeba na základě historie stavů vnějšího prostředí predikovat jeho budoucí stav. Jedná se de facto o konekcionistickou implementaci výsledků Husserlovy fenomenologické analýzy časového vědomí, které se v každém okamžiku skládá nejen z aktuálních dojmů, které jsou založeny na bezprostředně dostupných podnětech z vnějšího prostředí, ale také z implicitní paměťové stopy toho, co přítomnému okamžiku předcházelo - tzv. *retence* -, a z očekávání, co bude následovat - tzv. *protence* (Husserl, 1996). Vzhledem k tomu, že anticipace a schopnost zohledňovat předchozí zkušenost je součástí většiny kognitivních funkcí, představuje SRN jeden z nejužitečnějších



nástrojů konekcionistického modelování kognitivních procesů. Jako příklad zde může posloužit využití SRN k modelování schopnosti dítěte se na základě zkušenosti s

Obrázek 143: Ilustrace povahy „časového vědomí“ jednoduché rekurentní sítě, jejíž skrytá vrstva ve svém aktuálním stavu vždy zrcadlí svůj předešlý stav, který zase odráží jemu předcházející stav atd. V konečném důsledku tak jednoduchá rekurentní síť ve svém aktuálním stavu vždy implicitně nese stopu veškeré předcházející historie svých předešlých stavů.

řečovým projevem dospělých naučit rozlišovat hranice jednotlivých slov ve spojitém toku řeči. Elman (1990) takto prokázal, že SRN lze naučit segmentovat spojitý proud hlásek na jednotlivá slova čistě jenom na základě statistických zákonitostí v souvýskytu jednotlivých hlásek v řečovém projevu anglického mluvčího.

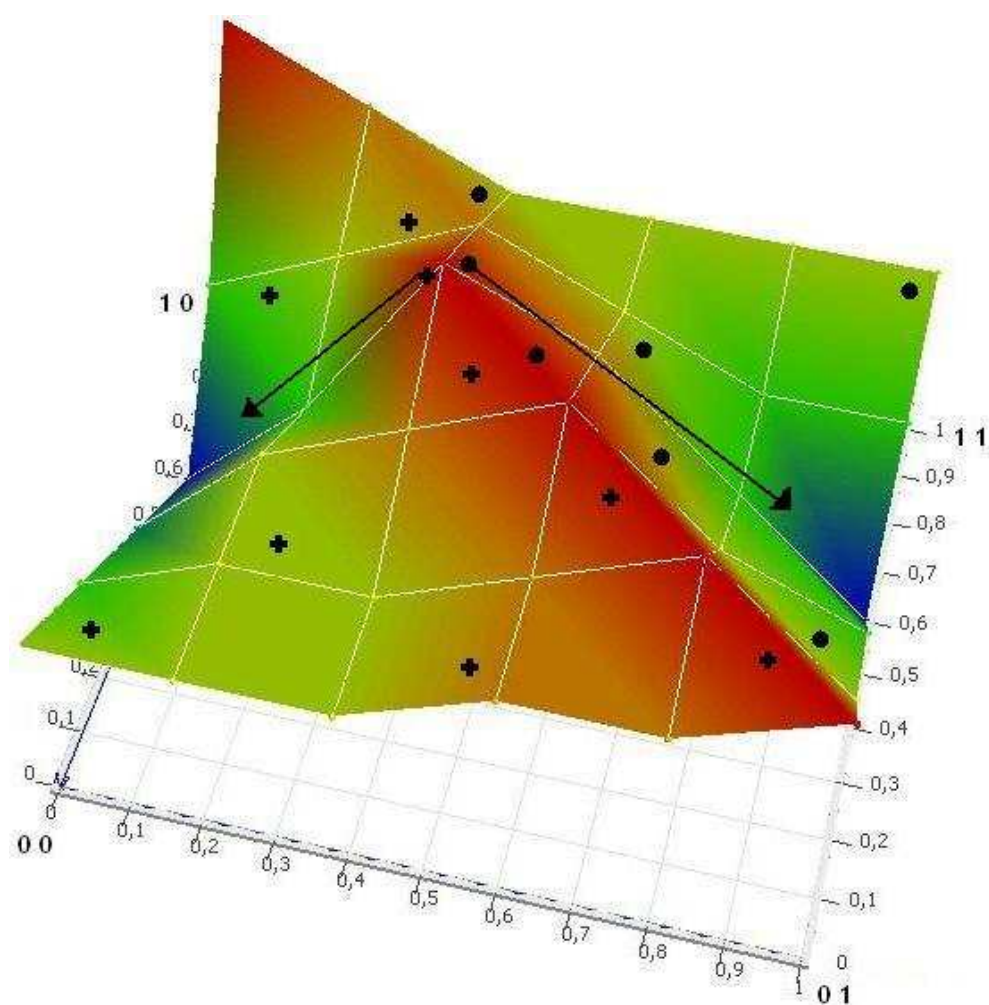
2.2.2.3.3 Atraktorové sítě

Stejně jako v případě SRN charakterizuje konektivitu atraktorové sítě přítomnost zpětnovazebních spojů. Na rozdíl od SRN se však tyto zpětnovazební spoje nacházejí mezi uzly z výstupní vrstvy a uzly z některé z předcházejících vrstev, tedy buď ze vstupní nebo skryté vrstvy. Tento způsob zapojení umožňuje výstupním uzlům měnit v čase (chápaném zde jako posloupnost výpočetních cyklů aktualizace míry aktivace jednotlivých uzlů sítě) míru své aktivity do té doby, dokud se tato nějak neusadí a nestabilizuje. Chování sítě s takovou konektivitou lze tedy charakterizovat jako postupné usazování se jejího stavu - definovaného mírou aktivace výstupních uzlů - v jednom z konečné množiny stabilních atraktorů, které jsou výstupem a odpovědí neuronové sítě na prezentovaný podnět, který se nachází v basin of attraction některého z atraktorů neuronové sítě.

Z hlediska konekcionistického modelování kognitivních funkcí se jako důležitá vlastnost atraktorové sítě jeví být její schopnost modelovat reakční časy v podobě počtu výpočetních cyklů nutných k usazení sítě v některém z konečné množiny atraktorů v reakci na prezentovaný podnět. Druhou užitečnou vlastností atraktorových sítí je jejich schopnost realizovat jakoukoli arbitrární transformaci vstupů na výstupy, která je zcela nezávislá na povrchní ne/podobnosti mezi vstupními podněty. Jinými slovy, atraktorová síť prostřednictvím specifického nastavení své váhové matice a prahových hodnot jednotlivých uzlů dokáže vytvarovat krajinu atraktorů a basins of attraction způsobem, který zajistí, že i velice nepodobné a ve vstupním reprezentačním prostoru vzdálené podněty nakonec skončí ve stejném atraktoru (i když ne nutně ve stejném čase), a naopak, tedy že podobné a ve vstupním reprezentačním prostoru blízké podněty nakonec skončí v odlišných atraktorech. Jedná se de facto pouze o jiný způsob realizace „reprezentační gravitační čočky“, která deformuje vzdálenosti mezi reprezentacemi podnětů ve vstupní vrstvě. Na rozdíl od statických vnitřních reprezentací zajišťovaných uzly ve skryté vrstvě se zde k dosažení stejného efektu využívá dynamického chování neuronové sítě v čase.

Na obrázku 144 je tento mechanismus schematicky znázorněn pomocí grafu stavového prostoru jednoduché atraktorové sítě, jejíž úloha spočívá v klasifikaci dvoubitových vstupních

signálů do dvou různých kategorií. Možné vstupy jsou v grafu reprezentovány 2-D plochou ohraničenou body (0 0), (1 0), (0 1), (1 1). Stejná plocha odpovídá také možným výstupům. Transformaci vstupů na výstupy v atraktorové síti je pak možné chápat jako postupný pohyb sítě z pozice, které odpovídá vstupu, na místo, které odpovídá požadovanému výstupu. Aby se síť nakonec dostala na požadovanou pozici, je nutné, aby stavový prostor sítě měl určitý specifický „výškový profil“ či „zvlnění“ odpovídající krajině atraktorů a jejich basins of attraction. V grafu tomuto „výškovému profilu“ odpovídá jeho třetí rozměr. Z obrázku jsou celkem dobře patrné dva atraktory („modré nížiny“) a jejich basins of attraction (okolní „svahy“), které stahují jednotlivé vstupy k danému atraktoru (výstupu). Křížkem a tečkou jsou pak v grafu zachyceny dvě skupiny vstupních podnětů, které se má neuronová síť naučit klasifikovat do dvou různých kategorií. Díky specifickému „zvlnění“ stavového prostoru - danému nastavením váhové matice a prahových hodnot jednotlivých uzlů – se všechny vstupy v různém počtu výpočetních cyklů nakonec přetransformují do stavu sítě, který odpovídá požadovanému výstupu. Šipkami jsou naznačeny vývojové trajektorie neuronové sítě po



Obrázek 144: Stavový prostor atraktorové sítě (více viz hlavní text).

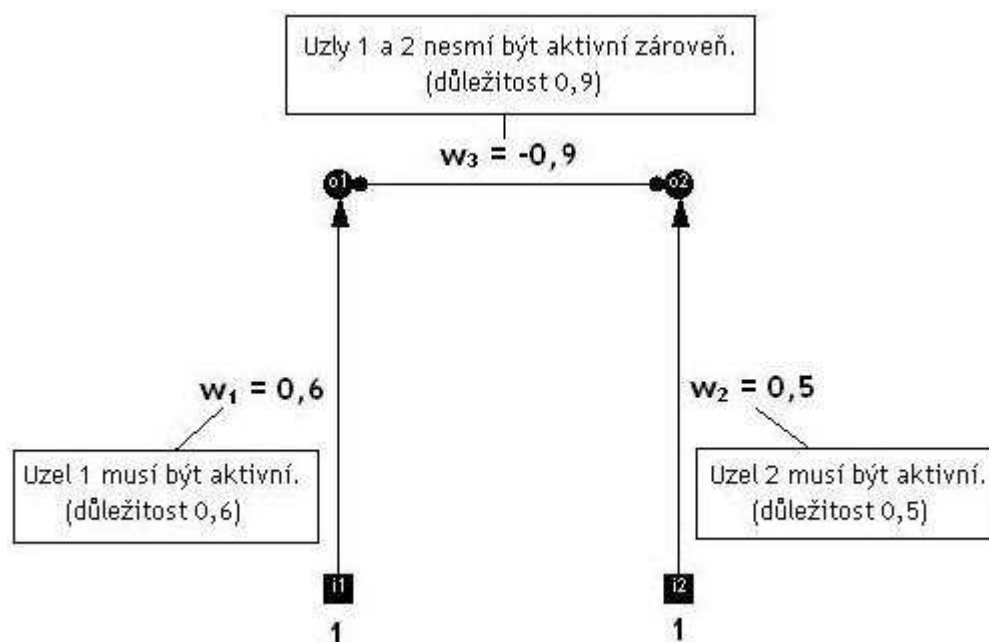
prezentaci dvou podnětů, které jsou ve vstupní vrstvě neuronové sítě reprezentovány jako dva sobě velice podobné, téměř identické vstupní aktivační vzorce; hranice mezi basins of attraction dvou atraktorů, která se táhne napříč stavovým prostorem, však umožňuje, aby tyto dva na první pohled velice podobné podněty byly klasifikovány do dvou různých kategorií. V grafu lze vidět i opačný případ, kdy se dva podněty, které jsou ve vstupní vrstvě kódovány jako rozdílné a sobě nepodobné, nacházejí v basin of attraction stejného atraktoru, tzn. že i přes svou povrchní rozdílnost jsou neuronovou sítí „chápány“ jako podněty, které si jsou na nějaké hlubší či abstraktnější rovině podobné. Jako konkrétní příklad uplatnění arbitrárního mapování vstupů na výstupy v rámci nějaké kognitivní funkce lze uvést osvojování si souvislostí mezi sémantickou a fonetickou stránkou řeči: Například slova „pes“ a „hafan“ odkazují ke stejnému nebo velice podobnému významu a to přesto, že z fonetického hlediska se jedná o dvě naprosto odlišná slova, která spolu nesdílejí ani jednu hlásku (McLeod, Plunkett, Rolls, 1998).

Způsob zpracovávání informací atraktorovou sítí se často charakterizuje jako druh optimalizace či harmonizace stavu neuronové sítě, která spočívá v hledání (počítání) takových hodnot aktivace skrytých a výstupních uzlů, které společně s danými hodnotami aktivace vstupních jednotek v maximální možné míře respektují všechna „měkká“ omezení řešené úlohy¹, která jsou zakódována do vážených spojů mezi jednotlivými uzly sítě. Stav maximální harmonie odpovídá atraktorům neuronové sítě - stabilním stavům, ve kterých neuronová síť končí dříve nebo později poté, co jí jsou prezentovány určité vstupní aktivační vektory. Jsou to stavy, kdy se již nemění míra aktivace žádného ze skrytých nebo výstupních uzlů, neboť jakákoli další změna v míře jejich aktivace by vedla jen ke snížení míry uspokojení jednotlivých omezení. Jakkoli velká je taková harmonizace, téměř nikdy nejsou plně uspokojena všechna omezení, neboť ta jsou často protichůdná a působí proti sobě. Výsledný stav sítě je tak vždy určitým kompromisem, ovšem kompromisem mezi různě „silnými hráči“: To, v jaké míře je které omezení nakonec respektováno, je dáno jeho důležitostí a ta je zakódována do váhy odpovídajícího spoje.

Paul Smolensky a Geraldine Legendre (2006, s. 15-17) celý proces harmonizace ilustrují na jednoduché atraktorové síti (viz obrázek 145), která se skládá ze dvou vstupních uzlů, které

¹ Na rozdíl od „tvrdých“ omezení „měkká“ omezení (*soft constraints*) nějaké úlohy nekorespondují přímočaře s pojmy, kterými je v rámci běžného diskurzu taková úloha většinou charakterizována. Jinými slovy, „tvrdá“ omezení mají sémantiku odpovídající běžnému každodennímu diskurzu. Příkladem by zde mohlo být jedno z pravidel popisujících proces usuzování na emocionální stav člověka na základě výrazu jeho tváře: „*Je-li na tváři člověka úsměv, potom je tento člověk šťastný*“. V případě „měkkých“ omezení je jejich sémantika striktně omezena pouze na vnitřní fungování neuronové sítě. Přeložíme-li takové „měkké“ omezení do běžného jazyka, mohlo by mít například podobu pravidla „*Je-li uzel 1 aktivní, musí být aktivní také uzel 2*“ (k tomuto tématu viz také s. 109-111).

mají pevně fixovanou míru aktivace na hodnotě 1 a z nichž každý prostřednictvím pozitivního váženého spoje vysílá aktivační signál k jednomu ze dvou uzlů ve výstupní vrstvě, které používají jednoduchou lineární aktivační funkci. Váha spoje do prvního výstupního uzlu (w_1) má hodnotu 0,6; váha spoje do druhého výstupního uzlu (w_2) pak 0,5. Mezi výstupními uzly je ještě jeden reciproční inhibiční spoj (w_3) s váhou -0,9, který způsobuje, že aktivita jednoho výstupního uzlu potlačuje aktivitu toho druhého. Máme zde tedy tři „měkká“ omezení, která by bylo možné přeložit do běžného jazyka jako tři pravidla, kterými by se mělo řídit chování neuronové sítě po prezentaci daného vstupního podnětu: 1) „Výstupní uzel 1 musí být aktivní“ (w_1) 2) „Výstupní uzel 2 musí být aktivní“ (w_2) a 3) „Výstupní uzel 1 a 2 nesmí být aktivní zároveň“ (w_3). Na první pohled je zřejmé, že tato pravidla jsou v rozporu a že nemohou být uspokojena všechna zároveň. Je tedy nutný určitý kompromis. Zde přichází ke slovu



Obrázek 145: Jednoduchá atraktorová síť s verbalizovanými omezeními, která jsou zakódována do vah jednotlivých spojů mezi uzly sítě. (s úpravami převzato z Smolensky, Legendre, 2006)

důležitost jednotlivých omezení, která jsou zakódována do vah odpovídajících spojů. Nejsilnější je třetí omezení (-0,9), tzn. že aktivní může být pouze jeden z výstupních uzlů. Nyní je otázka, který ze dvou výstupních uzlů zůstane aktivní a který svou aktivitu utlumí; jinými slovy, které ze dvou zbývajících omezení bude uspokojeno: Mohlo by to být druhé omezení, to by však nebylo optimálním řešením, neboť jeho důležitost je menší (0,5) než důležitost prvního omezení (0,6). Necháme-li tedy neuronovou síť zpracovat vstupní aktivační vzorec (1 1), dříve či později se ocitne ve stavu, kdy výstupní uzel 1 bude aktivní a

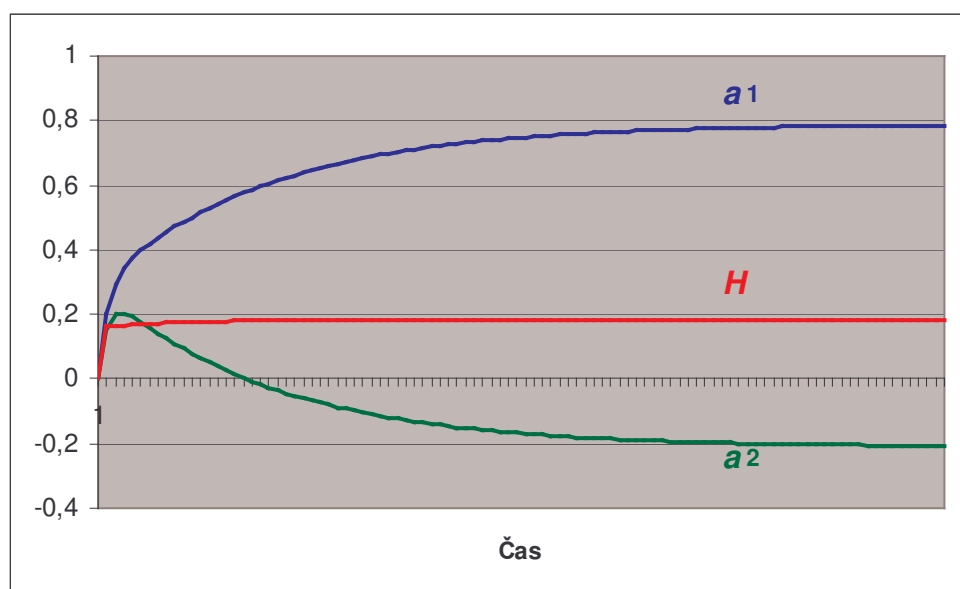
výstupní uzel 2 bude „mlčet“. V grafu 14 zachycujícím dynamiku chování sítě, která je dána dvojicí diferenciálních rovnic

$$\frac{da_1}{dt} = -a_1 + w_1 - w_3 a_2 \quad \frac{da_2}{dt} = -a_2 + w_2 - w_3 a_1,$$

je dobře vidět, jak míra aktivace uzlu 1 a 2 (a_1 , a_2) postupně směřuje k hodnotě 0,79, resp. -0,21, a jak odpovídající míra harmonie (H) počítaná podle vzorce

$$H(a) = \sum_{ij} a_i w_{ij} a_j - \sum_i a_i^2$$

postupně stoupá ke svému maximu. Tento stav neuronové sítě, kdy hodnota míry aktivace jejích dvou výstupních uzlů je 0,79 a -0,21, představuje při daném vstupu (1 1) a použité

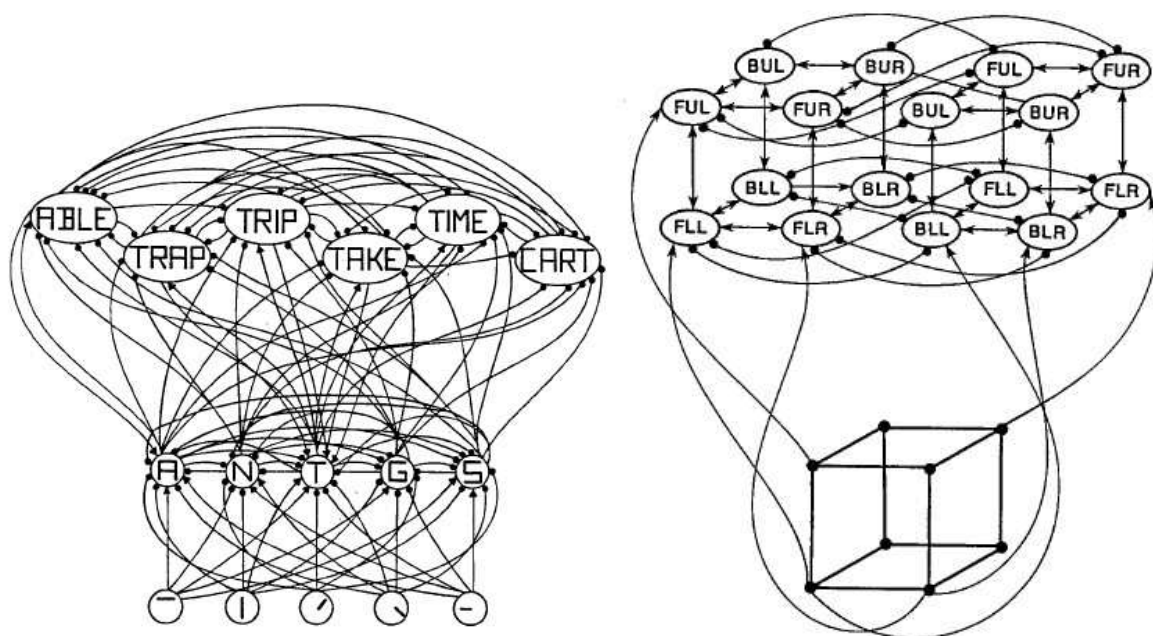


Graf 14: Vývoj míry aktivace dvou výstupních uzlů a_1 a a_2 a vývoj odpovídající míry harmonie H . (s úpravami převzato z Smolensky, Legendre, 2006)

aktivační funkci optimální kompromisní řešení, kdy jsou v maximální možné míře uspokojena tři konfliktní omezení implementovaná ve váhové matici neuronové sítě.

Příklad rozsáhlejší atraktorové sítě je možné vidět na obrázku 38 (s. 92), kde je zobrazen jednoduchý konekcionistický model asociační paměti v podobě série čtyř „momentek“ zachycujících proces postupného usazování se aktivačního vzorce uzlů sítě do té podoby, ve které nejlépe vyhovuje jednotlivým omezením v podobě dotazového (vstupního) aktivačního vzorce a v podobě vah jednotlivých spojů, ve kterých jsou zakódovány všechny znalosti

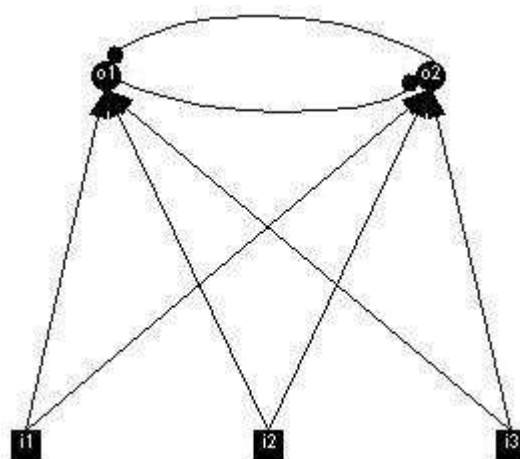
neuronové sítě o deseti různých domácích zvířatech. Jiným příkladem atraktorové sítě je známý McClellandův a Rumelhartův interaktivně-aktivační model rozpoznávání psaných slov, který se skládá ze tří rekurentně propojených vrstev uzlů odpovídajících vizuálnímu zpracovávání psaného textu 1) na úrovni dílčích znaků jednotlivých písmen, 2) na úrovni celých písmen a 3) na úrovni celých slov (viz obrázek 146a). Proces rozpoznávání jednotlivých slov je zde modelován jako proces hledání takového aktivačního vzorce, který by vzhledem k informacím přijatým ve vstupní vrstvě v maximální možné míře uspokojoval jednotlivá omezení vtělená do vah spojů. Identifikace slova pak odpovídá usazení se neuronové sítě v jednom z konečné množiny pevných bodů/atraktorů. Z dílny stejných autorů (Rumelhart a kol., 1986a) pochází také rekurentní síť modelující proces vnímání percepčně neurčitého či ambivalentního objektu v podobě průhledné, tzv. Neckerovy krychle, kterou můžeme vnímat buď shora, nebo zdola, ale nikdy ne oběma způsoby najednou (viz obrázek 146b). I zde neuronová síť hledá stav maximální možné harmonie, který odpovídá jednomu ze dvou možných způsobů vnímání Neckerovy krychle. To, ve kterém ze dvou atraktorů neuronová síť nakonec skončí, je ovlivněno náhodnými fluktuacemi v míře aktivace jednotlivých uzlů nebo simulovanou „únavou“ (tzv. akomodací) aktivních uzlů. Tyto změny umožňují cyklické přecházení mezi oběma atraktory, což odpovídá spontánnímu překlápění dvou způsobů interpretace Neckerovy krychle pozorovanému u lidí.



Obrázek 146: a) McClellandův a Rumelhartův interaktivně-aktivační model rozpoznávání slov. b) Konekcionistický model vnímání percepčně ambivalentní Neckerovy krychle (Vysvětlivky: F – front, přední, B – back, zadní, U – upper, horní, L – lower, dolní, R – right, napravo, L – left, nalevo)

2.2.2.3.4 Kompetitivní síť

Druhem atraktorové či rekurentní sítě jsou také tzv. *kompetitivní síť* (*competitive networks*). Jejich specifikem je to, že všechny jejich zpětnovazební spoje jsou pouze inhibiční, všechny mají stejnou a pevně fixovanou váhu a navíc navzájem spojují pouze uzly ve výstupní vrstvě (viz obrázek 147). Ve svém důsledku to vede k tomu, že jednotlivé výstupní uzly mezi sebou „bojují“ o to, kdo bude moci zareagovat svou aktivitou na prezentovaný vstupní aktivační vzorec: Ve vstupní vrstvě je nejdříve neuronové síti prezentován vstupní aktivační vzorec; to vede k šíření aktivačního signálu do výstupní vrstvy, kde se jednotlivé uzly v různé míře aktivují; vzhledem k tomu, že mezi výstupními uzly jsou reciproční inhibiční spoje se stejnou vahou, uzel, který je nejvíce aktivován, nakonec „umlčí“



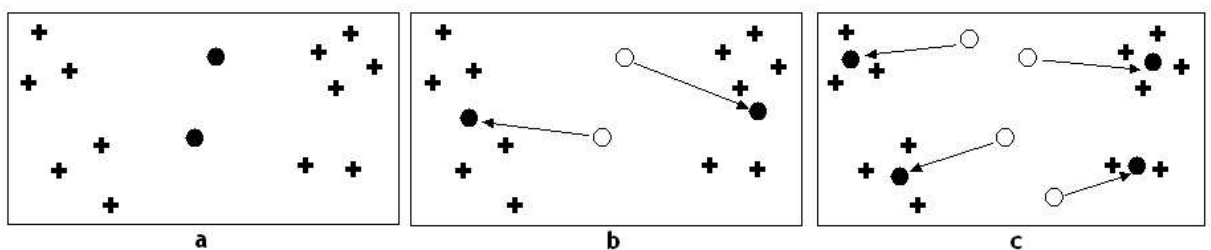
Obrázek 147: Kompetitivní síť. (vytvořeno s pomocí simulačního softwaru Tlearn)

aktivitu všech ostatních uzlů ve výstupní vrstvě; celý „souboj“ má tedy pouze jednoho vítěze. To, kdo zvítězí, je samozřejmě dáno kombinací vstupního aktivačního vektoru a nastavení váhové matice, která determinuje, kolik aktivačního signálu obdrží jednotlivé uzly z výstupní vrstvy. Dříve než se kompetitivní síť naučí něco užitečného, je vítězství konkrétního výstupního uzlu čistě náhodné, neboť náhodné je také výchozí nastavení váhové matice neuronové sítě. Bez ohledu na to, zda je vítězství dílem náhody nebo „vlastního přičinění“, kompetitivní síť může takovou informaci o vítězi využít jako první vstup do tzv. **kompetitivního učícího algoritmu** (*competitive learning algorithm*), jehož podstata spočívá v úpravě vah pouze těch spojů, které vedou z uzlů ve vstupní vrstvě do vítězného výstupního uzlu. Pravidlo pro výpočet úprav vážených spojů má podobu následujícího vzorce:

$$\Delta w_{ij} = 0 \text{ jestliže uzel } i \text{ prohraje}$$

$$\Delta w_{ij} = \varepsilon (a_j - w_{ij}) \text{ jestliže uzel } i \text{ vyhraje}^1$$

kde ε je obvyklý parametr míry učení, který specifikuje velikost korekce vah spojů, a_j je míra aktivace vstupního uzlu, z něž vede spoj do vítězného výstupního uzlu a w_{ij} je váha tohoto spoje. Z rovnice je dobře patrné, že výsledkem opakované aplikace tohoto učícího pravidla je postupné přibližování hodnoty váhy spoje (w_{ij}) hodnotě míry aktivace odpovídajícího vstupního uzlu j (a_j): Je-li a_j větší než w_{ij} , vede to ke zvýšení váhy spoje w_{ij} ; je-li a_j menší než w_{ij} , korekce má podobu snížení váhy spoje w_{ij} . Váhový vektor vítězného uzlu se takto začne čím dál tím více podobat odpovídajícímu aktivačnímu vektoru ve vstupní vrstvě. A čím jsou si dva vektory podobnější, tím větší je jejich tzv. *dot product*, který je sumou násobků čísel z odpovídajících pozic ve dvou vektorech.¹ A čím větší je dot product, tím větší je míra aktivace vítězného uzlu. Takže při příští prezentaci stejného nebo podobného vstupního aktivačního vzorce je vysoká pravděpodobnost, že se znovu nejvíce aktivuje původní uzel a že tento uzel znovu zvítězí v konkurenci ostatních uzlů ve výstupní vrstvě. Tímto způsobem lze kompetitivní síť naučit rozlišovat v souboru vstupních podnětů různé kategorie objektů, jejichž počet odpovídá počtu uzlů ve výstupní vrstvě. Na obrázku 148a je zachycena struktura podobnosti dvanácti vstupních aktivačních vektorů (křížky) a dvou váhových vektorů dvou výstupních uzlů (černé tečky). Před započítím kompetitivního učení je pozice váhových vektorů čistě náhodná. Po několika kolech učení se však váhové vektory změní tak, že každý z nich se podobá jedné podmnožině vstupních aktivačních vektorů (obrázek 148b).



Obrázek 148: Struktura podobnosti dvanácti vstupních aktivačních vektorů (křížky) a dvou, resp. čtyř váhových vektorů (černé tečky). a) Stav před započítím kompetitivního učení, kdy je pozice váhových vektorů ve vztahu ke vstupním aktivačním vektorům zcela náhodná. b) Situace po několika kolech kompetitivního učení, kdy se váhové vektory již podobají dvěma podmnožinám vstupních aktivačních vektorů. Tento stav odpovídá schopnosti kompetitivní sítě vstupní podněty klasifikovat do dvou velice obecných kategorií. c) Situace po přidání dalších dvou uzlů do výstupní vrstvy: jednotlivé váhové vektory se nyní mnohem těsněji podobají čtyřem podmnožinám prezentovaných podnětů, které odpovídají čtyřem specifitějším kategoriím. (s úpravami převzato z McLeod, Plunkett, Rolls, 1998, s. 131)

¹ Například dot product dvou vektorů $[1 \ 2 \ 1 \ 0]$ a $[0 \ 3 \ 1 \ 4]$ se rovná $((1 \times 0) + (2 \times 3) + (1 \times 1) + (0 \times 4)) = 7$.

Kompetitivní síť se takto naučila v prezentovaných podnětech rozlišovat dvě velice obecné kategorie objektů – cosi na způsob taxonu „primátů“ nebo „šelem“. Když se do výstupní vrstvy přidají další dva uzly, kompetitivní síť tím získá možnost upravit váhové vektory tak, aby se mnohem těsněji podobaly čtyřem různým podmnožinám vstupních aktivačních vektorů (obrázek 148c). Výsledkem jsou čtyři specifitější kategorie – typu „psovitých a kočkovitých šelem“ nebo „poloopic“ a „vyšších primátů“, použijeme-li opět příklady z biologické taxonomie. Jemnost klasifikace, typy a počet kategorií, které kompetitivní učící algoritmus ve vstupech objeví, takto kriticky závisí na počtu uzlů ve výstupní vrstvě neuronové sítě.

Celý proces kompetitivního učení je de facto formou shlukové či faktorové analýzy, která vstupní data s podobnými vlastnostmi soustřeďuje do shluků, jejichž počet odpovídá počtu uzlů, které má kompetitivní síť ve výstupní vrstvě k dispozici. Kompetitivní síť lze takto použít k redukci velikosti vstupních vektorů jejich mapováním do menšího počtu shluků či faktorů. V rámci konekcionistického modelování kognitivních funkcí se této schopnosti kompetitivní sítě často využívá k předzpracovávání různých vstupních podnětů. Takto zpracované podněty pak slouží jako vstup do dalšího typu neuronové sítě, například do běžné dopředné neuronové sítě trénované již s pomocí tradičních učících algoritmů jako je Hebbovo pravidlo nebo delta pravidlo. Toto předzpracování přitom může mít podobu kategorizace nebo „ortogonizace“: Při kategorizaci dochází k odstranění redundance obsažené ve vstupu tím, že se několika podobným vstupním vzorcům přisoudí jeden výstupní uzel; po zpracování vstupních podnětů tak výstupy mezi sebou korelují více než původní vstupy. Opačný proces nastává v případě ortogonizace, kdy částečně korelujícím podnětům jsou přisouzeny výstupy, které spolu korelují mnohem méně než původní vstupy. „Ortogonizace“ nachází své využití zejména při předzpracovávání vstupů pro asociátor vzorců (*pattern associator*), který do své váhové matice dokáže uložit tím více asociací, čím méně jsou si jednotlivé vstupní vzorce podobné, tedy čím méně spolu korelují (viz také s. 220-223). To, jestli bude kompetitivní síť kategorizovat nebo „ortogonizovat“ přitom závisí na počtu výstupních uzlů a vstupních podnětů (McLeod, Plunkett, Rolls, 1998).

Z hlediska možností konekcionistického modelování kognitivních funkcí je důležitou vlastností kompetitivního učení také to, že na rozdíl od Hebbova pravidla nebo delta učícího algoritmu nepotřebuje žádnou zpětnou vazbu od pomyslného „vševědoucího“ učitele, který by neuronové síti říkal, jaký má být správný výstup a jakým způsobem je tedy potřeba upravit stávající nastavení váhové matice. Kompetitivní síť si v tomto ohledu vystačí zcela sama, neboť to jediné co potřebuje zjistit, aby mohla „nastartovat“ kompetitivní učící algoritmus, je který z výstupních uzlů zvítězil v „souboji“ o možnost se aktivovat v reakci na prezentovaný

podnět. Kompetitivní síť tak patří do širší skupiny tzv. samotrénujících se či samoorganizujících se sítí (*self-organizing networks*), které používají učící algoritmy bez supervidujícího „učitele“ (*unsupervised learning algorithms*). Kompetitivní síť lze tedy používat jako součást konekcionistických modelů kognitivních funkcí i tam, kde víme, že lidský mozek nedostává žádnou zpětnou vazbu využitelnou při úpravě vah synaptických spojů. „Sympatická“ je rovněž biologická plausibilita kompetitivních sítí, jejichž reciproční inhibiční konektivitu lze nalézt na mnoha místech v lidském mozku.

2.2.2.3.5 Autoasociátory

Specifickým druhem rekurentní sítě je autoasociátor, který se skládá pouze z jedné vrstvy uzlů, které jsou prostřednictvím modifikovatelných zpětnovazebních spojů napojeny na vstup všech uzlů sítě (kromě svého vlastního vstupu). Celkový vstup do uzlu i ($netinput_i$) je tak tedy sumou vnějšího vstupu ($extinput_i$) a vnitřního vstupu, který je součtem násobků míry aktivity uzlů propojených s uzlem i a vah spojů mezi nimi:

$$netinput_i = extinput_i + \sum_j a_j w_{ij}$$

Vnější vstup představuje druh senzorické stimulace, která v závislosti na použité aktivační funkci vede k určité míře aktivace jednotlivých uzlů. Tato aktivace se pak po vynásobení vahou odpovídajícího spoje stává vnitřní komponentou celkového vstupu do dalšího výpočetního cyklu. Tyto výpočetní cykly se opakují do té doby, dokud autoasociátor nedosáhne stabilního stavu, kdy se hodnota vnitřního vstupu mezi jednotlivými výpočetními cykly již nemění. Riziko nekontrolovaného růstu míry aktivace v důsledku možné pozitivní zpětné vazby je zde eliminováno použitím nelineární aktivační funkce, která povoluje aktivaci pouze v určitém omezeném rozmezí, konkrétně mezi -1 a 1 . Toto konkrétní rozmezí autoasociátoru umožňuje používat určitou specifickou verzi delta učícího algoritmu a následně tak i plnit svou funkci, která – jak již napovídá samotný název autoasociátoru – spočívá v asociování dvou stejných vstupních aktivačních vzorců. Jinými slovy, autoasociátor na výstupu reprodukuje stejný vzorec aktivity, který je mu prezentován na vstupu. Autoasociátor si tyto asociace osvojuje s pomocí delta učícího algoritmu, který používá chybový signál spočívající v rozdílu mezi vnějším a vnitřním vstupem:

$$\delta_i = extinput_i - \sum_j a_j w_{ij} .$$

Korekce vah zpětnovazebních spojů se počítá podle vzorce

$$\Delta w = \varepsilon \delta_i a_j,$$

kde ε je parametr míry učení, δ_i je rozdíl mezi vnějším a vnitřním vstupem do uzlu i a a_j je míra aktivace vysílajícího uzlu j , která může nabývat hodnot -1 až 1 . Cílem procesu učení je, aby se rozdíl mezi vnějším a vnitřním vstupem rovnal nule. Cesta k tomuto cíli vede přes posilování vah těch spojů, které tento rozdíl redukuje, a oslabování vah těch spojů, které k rozdílu naopak přispívají: Autoasociátoru se nejdříve prezentuje vstupní podnět a aktivační signál se nechá sítí kolovat, dokud se síť neusadí v nějakém stabilním stavu; když je vnější vstup větší než vnitřní vstup, učicí algoritmus posiluje váhy těch spojů, které přinášejí pozitivní vstup, a oslabuje váhy těch spojů, které přinášejí negativní vstup; při následující prezentaci vstupního podnětu se tak zvýší vnitřní vstup a rozdíl mezi vnějším a vnitřním vstupem se zredukuje; je-li naopak vnější vstup menší než vnitřní vstup, jsou oslabovány váhy spojů přenášejících pozitivní vstup a zesilovány váhy spojů zprostředkovávajících negativní vstup; při následující prezentaci vstupního podnětu se tak sníží vnitřní vstup a rozdíl mezi vnějším a vnitřním vstupem se zmenší. Jakmile se δ_i rovná nule, proces učení se automaticky zastaví.

Z popisu tohoto učicího algoritmu by mělo být zřejmé, že výsledná váhová matice je de facto maticí korelační, která udává, jakým způsobem (+/-) a v jaké míře spolu aktivita jednotlivých uzlů sítě koreluje: Cílem učení je, aby se vnější vstup rovnal vstupu vnitřnímu. Jestliže vnější vstup do jednoho z uzlů je $+1$, potom učicí algoritmus nastaví váhy spojů vedoucích z dalších uzlů tak, aby také vnitřní vstup do daného uzlu byl pozitivní; díky tomu, že míra aktivace se může pohybovat v rozmezí -1 až 1 , jednotlivé váhy tak budou mít povahu korelačních koeficientů vyjadřujících míru korelace mezi aktivitou jednotlivých uzlů. Ve váhové matici autoasociátoru, který se naučil na výstupu reprodukovat vstupní aktivační vzorec $[1 \ -1 \ 1 \ -1]$, takto bude zakódováno, že aktivita prvního uzlu pozitivně koreluje s aktivitou druhého uzlu a negativně s aktivitou druhého a čtvrtého uzlu. Převáděno do řeči vah to znamená, že první uzel bude s třetím uzlem propojen excitačním spojením a s druhým a třetím uzlem pak spojením inhibičním – toto konkrétní nastavení vah zapříčiní pozitivní vnitřní vstup do prvního uzlu a minimalizuje rozdíl mezi vnějším a vnitřním vstupem. V případě druhého uzlu, jehož vnější vstup je -1 , je pak nastavení vah přesně opačné. Důležité je, že stejně jako je první uzel propojen excitačně s třetím uzlem a inhibičně s druhým uzlem, je také třetí, resp. druhý uzel excitačním, resp. inhibičním spojením propojen s prvním uzlem.

Takto vzniká cosi na způsob velice komplexního podpůrného systému, kde se jednotlivé uzly navzájem podepírají a podporují ve snaze reprodukovat určitý vstupní aktivační vzorec.

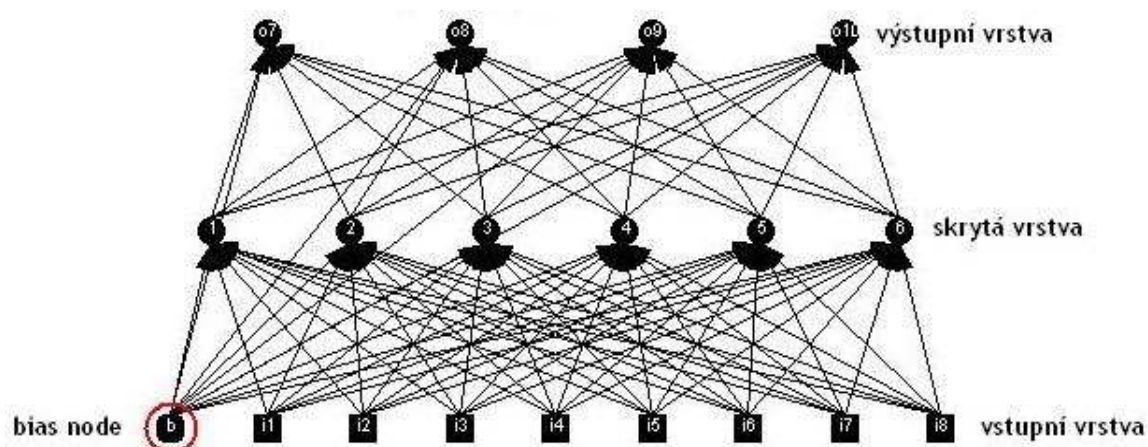
Díky tomu, že autoasociátor má ve své váhové matici zakódované informace o tom, jak spolu koreluje aktivita všech uzlů sítě, autoasociátor může tyto informace využít k reprodukci původního aktivačního vzorce i na základě jen neúplných a nebo nějak poškozených vstupních signálů (*pattern completion*): Má-li na jedné straně autoasociátor k dispozici alespoň určitou minimální část vstupního aktivačního vektoru - tzn. má-li k dispozici informaci o tom, jaká je (správná) míra aktivace určitého minimálního počtu uzlů ve vstupní vrstvě - a jestliže na straně druhé autoasociátor disponuje informacemi o tom, jak aktivita těchto uzlů koreluje s aktivitou uzlů, jejichž míra aktivace je neznámá (nebo nějak zkreslená), potom má autoasociátor k dispozici všechny informace, které potřebuje k rekonstrukci původního aktivačního vzorce. Je to de facto jistá obdoba procesu replikace molekuly DNA, kdy zákonité párování bází (A-T, G-C) umožňuje doplnit komplementární části rozdělené molekuly DNA v průběhu jejího zdvojování. Podobnost lze vidět také v tom, že v obou případech je možnost rekonstrukce neúplné struktury dána určitými specifickými vlastnostmi vztahů mezi základními jednotkami daného systému - v případě kompletace molekuly DNA to jsou chemické vazby mezi čtyřmi dusíkatými bázemi a v případě rekonstrukce aktivačních vzorců to jsou vážené spoje mezi uzly sítě. Proces rekonstrukce neúplného vstupního vzorce je jinak opět procesem postupného usazování se neuronové sítě do stavu s maximální harmonií, kdy jsou při daném vstupu maximálně uspokojena všechna omezení implementovaná ve vahách spojů.

S pomocí autoasociátoru lze takto úspěšně modelovat (nejen) pro lidskou kognici charakteristickou schopnost vyvolávat z paměti kompletní informace na základě neúplných nebo informačním šumem zatížených nápověd (*recall cues*). Například během rozhovoru na „divokém“ večírku člověk obvykle dokáže správně identifikovat jednotlivá slova vyřčená jeho protějškem, přestože zvuky slov, které k němu doléhají, jsou značně zkreslené šumem pozadí nebo zvláštnostmi ve výslovnosti způsobené nadměrnou konzumací alkoholu. Odvrácenou stránkou této schopnosti je tendence autoasociátoru zaměňovat nové, ale známým vstupům podobné informace za neúplné vstupní aktivační vzorce, které je potřeba přetransformovat na ty známé a původní, což odpovídá obecné lidské tendenci interpretovat nové informace prizmatem předchozích zkušeností. Riziko této záměny lze zredukovat tím, že se sníží míra překrytí nového a známého aktivačního vzorce, tj. míra jejich korelace (tzv. „ortogonizace“). Toho lze dosáhnout například tím, že ještě před prezentací samotnému asociátoru se oba aktivační vzorce zpracují pomocí kompetitivní sítě (viz předchozí oddíl).

Podobně jako asociátor vzorců trénovaný s pomocí Hebbova pravidla (viz oddíl „2.2.2.2.1 Hebbovo pravidlo“), také autoasociátor dokáže do jedné váhové matice zakódovat větší počet různých asociací. Daleko zajímavější je schopnost autoasociátoru si čistě na základě své zkušenosti s jednotlivými příklady nějaké kategorie objektů osvojit jejího prototypického zástupce a to bez jakékoli explicitní instrukce či možnosti „vidět“ tento prototyp „na vlastní oči“. S pomocí takto osvojeného prototypu pak autoasociátor dokáže také správně kategorizovat nové a zcela neznámé objekty na základě jejich podobnosti s osvojeným prototypem. Ten přitom není ničím jiným než váhovou maticí, která díky superponování řady dílčích váhových maticí vznikajících při prezentaci jednotlivých příkladů dané kategorie kóduje informace (o korelaci aktivity jednotlivých uzlů), které jsou pravdivé pro většinu vstupních aktivačních vzorců reprezentujících jednotlivé konkrétní příklady dané kategorie. Tímto autoasociátor implementuje několik důležitých aspektů lidské kognice spočívajících ve spontánní a nesupervidované kategorizaci objektů, v tvorbě „neexistujících“ prototypů a ve schopnosti rozpoznávat nové a neznámé objekty jako příslušníky určité kategorie na základě jejich podobnosti s jejím prototypickým zástupcem.

2.2.2.3.6 Bias node

Na úplný závěr je zde potřeba ještě stručně zmínit tzv. *bias node* – speciální vstupní uzel, na který lze narazit v mnoha konekcionistických modelech. Tento uzel má konstantní míru aktivace, která je většinou nastavena na maximální možnou hodnotu +1 (v případě použití logistické aktivační funkce), a prostřednictvím běžných, učením modifikovatelných spojení ovlivňuje míru aktivace každého uzlu ve skryté a výstupní vrstvě (viz obrázek 149). Již



Obrázek 149: Třívrstevná neuronová síť se zobrazeným bias node (vlevo dole). (vytvořeno s pomocí simulačního softwaru Tlearn)

samotný název uzlu naznačuje povahu jeho funkce, která spočívá v usměrňování či vychylování (*biasing*) aktivity jednotlivých uzlů ve skryté a výstupní vrstvě. Směr a velikost takového vychýlení je determinována „nábojem“ a silou váženého spoje vedoucího z bias node do vybraného uzlu. V případě silného pozitivního spoje vstup z bias node nastavuje určitou bazální či spontánní míru aktivace uzlu, která je zcela nezávislá na aktuálním vstupu přicházejícím ze vstupní vrstvy neuronové sítě. Naopak v případě silného negativního spoje vstup z bias node funguje jako prahová hodnota, kterou je potřeba překonat dostatečně silnými signály ze vstupní vrstvy, než je vybranému uzlu umožněno se aktivovat.

2.3 Metodologie konekcionistických simulací kognitivních procesů

Konečným cílem každého vědeckého zkoumání nějakého jevu je jeho vysvětlení. Toto vysvětlení má podobu vědecké teorie, která představuje určité zevšeobecnující tvrzení o povaze příčinných a jiných vztahů mezi proměnnými, prostřednictvím kterých jsou popsány všechny jevy a události, o kterých se předpokládá, že se zkoumaným jevem nějak souvisí. Na základě takové teorie pak lze sestavit zjednodušený formální model zkoumaného jevu, který v sobě zachovává pouze ty jeho charakteristiky, o kterých se v dané teorii předpokládá, že jsou pro jeho vysvětlení důležité. Model nějakého jevu je tak de facto pouze detailnějším rozpracováním jeho teoretického vysvětlení do podoby zjednodušené verze zkoumaného jevu samotného. Srovnání chování formálního modelu s reálným chováním zkoumaného jevu pak umožňuje testovat teorii, ze které daný model vychází. Specifikum konekcionistických modelů spočívá v tom, že jsou implementovány v podobě počítačového programu. Počítačová simulace formálních modelů přitom s sebou přináší řadu výhod (Plunkett, Elman, 1997; O'Reilly, Munakata, 2000):

- 1) Nejčastěji uváděnou výhodou počítačových simulací je jejich **vysoká míra explicitnosti**: Když se člověk snaží nějaký teoretický model převést do podoby funkčního počítačového programu, nutí ho to k explicitní a precizní formulaci všech jeho teoretických předpokladů týkajících se chování zkoumaného jevu.
- 2) Člověku to umožňuje **se vyhnout nejrozumnějším homunkulům** – vágním psychologickým pojmům připisujícím různým částem kognitivního systému nejrozumnější kognitivní funkce, aniž by přitom skutečně vysvětlovaly, jak jsou tyto funkce realizovány (viz s. 48-49).
- 3) Člověka to také nutí detailně **zvažovat nové aspekty zkoumaného jevu**, které jsou při jeho pouze verbálním popisu prakticky „neviditelné“.

- 4) Explicitně formulovaný počítačový model člověku umožňuje generovat přesné **predikce důsledků vyplývajících z přijatých teoretických předpokladů**. Tato vlastnost počítačových modelů nabývá na důležitosti v případě komplexních modelů, které obsahují řadu nelineárních vztahů a závislostí, jejichž důsledky pro chování modelu lze jen velice obtížně předvídat.
- 5) Člověk takto může narazit na **nové a nepředvídané způsoby chování** simulačního modelu, které se následně může pokusit experimentálně ověřit i přímo na modelovaném systému. V případě umělých neuronových sítí „*může být mimořádně zajímavé a informativní určit, jaké zákonitosti sít' využívá. Jaké druhy zákonitostí ve vstupních vzorcích sít' objevila? Jak tyto zákonitosti reprezentuje? Jak jsou tyto zákonitosti kombinovány za účelem řízení reakce sítě? V mnoha případech odpovědi na tyto otázky mohou odhalit vlastnosti problémů a schémata pro reprezentování těchto vlastností, které by nikdo vůbec nepředvídal.*“¹ (Dawson, 2002)
- 6) Speciálně v případě konekcionistických výpočetních modelů má člověk **možnost pozorovat proces osvojování nějaké kognitivní funkce** v průběhu učebního procesu založeného na dostupných příkladech; stejně tak **lze sledovat ztrátu osvojené kognitivní funkce** v důsledku poškození funkční architektury neuronové sítě či **porovnávat míru poškození modelované kognitivní funkce** v závislosti na velikosti a umístění simulovaných lézí.
- 7) Počítačové modely člověku dávají **vysokou míru kontroly nad všemi důležitými proměnnými**, jejichž hodnotami lze snadno, přesně a cíleně manipulovat za účelem kauzální analýzy. V případě umělých neuronových sítí to znamená například možnost libovolně volit rozsah a umístění umělých lézí a následně sledovat jejich důsledky pro simulovanou kognitivní funkci, což je experimentální design, který lze v reálných podmínkách – z etických a technických důvodů - jen velice obtížně realizovat.²

¹ „... it can be extremely interesting, surprising, and informative to determine what regularities the network exploits. What kinds of regularities in the input patterns has the network discovered? How does it represent these regularities? How are these regularities combined to govern the response of the network? In many instances, the answers to these questions can reveal properties of problems, and schemes for representing these properties, that were completely unexpected.”

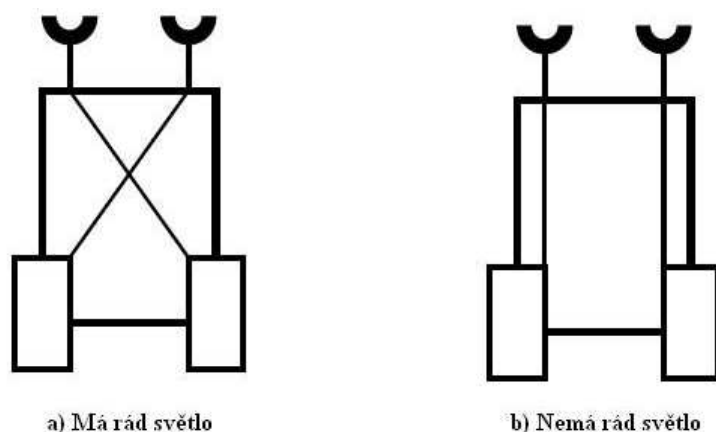
² Z výzkumných metod používaných v kognitivní neurovědě mají k tomuto experimentálnímu designu asi nejbližší *postmortální studie*, které spočívají v pečlivém zaznamenávání klinických příznaků pacientů s poškozeným mozkem v době, kdy žijí, a v lokalizaci lézí v jejich mozku po jejich smrti. Získané informace pak umožňují usuzovat na souvislost mezi poškozenou částí mozku a poškozenou duševní funkcí pozorovanou u pacienta. V poslední době se ke zkoumání vztahu mezi mozkem a chováním vedle různých funkčních zobrazovacích metod (*SPECT, PET, fMR*) používá také *repetitivní transkraniální magnetická stimulace* – neinvazivní metoda, která prostřednictvím silného magnetického pole umožňuje výzkumníkovi libovolně zapínat (excitovat) a vypínat (inhibovat) povrchové korové oblasti (zhruba 2x2x2 cm) lidského mozku.

- 8) S tímto úzce souvisí další výhoda počítačových simulací, která spočívá v tom, že člověku **umožňují přímo nahlédnou dovnitř (kauzálního) mechanismu odpovědného za chování modelovaného systému.** Tímto se metodologie počítačových simulací liší od tradičního výzkumného paradigmatu kognitivní psychologie, kde se na předpokládané charakteristiky vnitřního kognitivního mechanismu obvykle usuzuje pouze nepřímo na základě pozorování vnějších projevů nějaké kognitivní funkce a na základě testování předpovědí vyplývajících z předpokládaných vlastností kognitivního mechanismu. Počítačové simulace tedy výrazně napomáhají porozumění, proč se studovaný systém chová tak, jak se chová.

Často se přitom stává, že takové vysvětlení má značně protiintuitivní povahu. Například se může ukázat, že navenek velice složité chování může mít svou příčinu v jednoduchých lokálních interakcích mezi relativně jednoduchými funkčními komponentami systému. Německý kybernetik *Valentino Braitenberg* (1984) v této souvislosti hovoří o tzv. **zákonu vzestupné analýzy a sestupného vynalézání** (*the law of uphill analysis and downhill invention*), který formuloval na základě série myšlenkových experimentů s jednoduchými zařízeními, která jsou navzdory své jednoduché konstrukci schopna vykazovat překvapivě komplexní chování, při jehož vnějším pozorování má člověk automatickou tendenci (nesprávně) předpokládat, že musí být realizováno obdobně komplexním vnitřním řídicím mechanismem (viz obrázek 150). Braitenberg takto zjistil, že „*je... snadné vytvářet malé stroje, které dělají určité triky. Je také docela snadné pozorovat celý repertoár chování těchto strojů – a to i tehdy, když překračuje to, co jsme původně plánovali, jak se tomu často děje. Ale je mnohem obtížnější začínat z vnějšku a pokoušet se odhadnout vnitřní strukturu pouze na základě pozorovatelných údajů. [...] Analýza je mnohem obtížnější než vynalézání v tom slova smyslu, že indukce obecně člověku zabere více času než dedukce. Při indukci člověk musí po cestě pátrat, zatímco při dedukci člověk sleduje přímočarou cestu. Psychologický efekt toho je následující: když analyzujeme nějaký mechanismus, máme sklon přeceňovat jeho komplexitu.*“¹ (Braitenberg, 1984, s. 20-21)

Výsledkem aplikace tohoto zákona ve výzkumné praxi je tzv. **syntetická psychologie**, jejíž metodologii lze shrnout pod akronymem **SEA, syntéza-emergence-analýza**: První krok

¹ "It is... easy to create little machines that do certain tricks. It is also quite easy to observe the full repertoire of behavior of these machines – even if it goes beyond what we had originally planned, as it often does. But it is much more difficult to start from the outside and try to guess internal structure just from the observation of the data. [...] Analysis is more difficult than invention in the sense in which, generally, induction takes more time to perform than deduction: in induction one has to search for the way, whereas in deduction one follows a straightforward path. A psychological consequence of this is the following: when we analyze a mechanisms, we tend to overestimate its complexity"



Obrázek 150: Braitenbergovo Vozidlo 2. Komplexita chování inteligentního systému není nikdy čistě jenom funkcí jeho vnitřního řídicího mechanismu, ale také jeho senzory a aktuátory zprostředkované lokální interakce s vnějším prostředím (viz také Simonova „mravenčí metafora“ či Brooksovy roboty se subsumpční architekturou, s. 198-204). Braitenbergovo Vozidlo 2 tuto skutečnost demonstruje v praxi: Vozidlo 2 je vybaveno dvěma senzory a dvěma motory. Zatímco u Vozidla 2a jsou senzory s motory propojeny dvěma paralelními excitačními spoji, u Vozidla 2b jsou tyto excitační spoje překřížené. Senzory jsou přitom citlivé na světlo. Vozidlo 2b bude mít tendenci se vyhýbat veškerým světelným zdrojům, neboť senzor, který se dostane blíže ke zdroji světla, bude více excitován a v důsledku toho se také bude rychleji otáčet motor, který je blíže ke světelnému zdroji. Vozidlo 2a bude naopak díky svým překříženým excitačním spojem světelné zdroje vyhledávat. Relativně komplexní chování Braitenbergových vozidel je determinováno jednoduchým přímým propojením senzorů s aktuátory, a co je důležité, bez nutnosti nějakého mezilehlého kognitivního modulu, ve kterém by docházelo k symbolické re-reprezentaci okolí a aplikaci různých výpočetních operací na tuto symbolickou reprezentaci. (Na přiloženém CD je možné pod názvy „Lightfollow“ a „Lightavoid“ nalézt krátké videoukázky zachycující chování skutečných robotů postavených na základě myšlenkových experimentů obsažených v Braitenbergově knize „Vehicles: Experiments in Synthetic Psychology“; videoukázky byly ztaženy z osobních webových stránek kognitivního vědce Michaela R. W. Dawsona <http://www.bcp.psych.ualberta.ca/~mike/>.)

spočívá v kompletaci (syntéze) funkčního modelu ze souboru jeho stavebních prvků. Druhý krok spočívá v pozorování a zaznamenávání (emergentního) chování, které povstává z lokálních interakcí mezi stavebními prvky modelu. A poslední krok má pak podobu analýzy a vysvětlení vzniku vlastností emergentního chování modelovaného systému v termínech interakcí mezi jeho jednotlivými stavebními prvky (Dawson, 2004). Tato charakteristika metodologie syntetické psychologie platí do značné míry rovněž pro metodologii konekcionistického modelování kognitivních funkcí.

Počítačová simulace je de facto určitý druh experimentu a stejně jako běžný experiment i počítačová simulace by měla začít přesnou specifikací problému, který chceme vyřešit, a přesnou formulací otázky, na kterou chceme získat odpověď. Experimentátor by tedy měl mít jasně formulovanou hypotézu, kterou chce testovat, ještě před samotným sestavením

výpočetního (simulačního) modelu. Testování takové hypotézy může mít například podobu ověření předpovědi vyplývající z nějakého teoretického předpokladu o povaze určitého kognitivního procesu, nebo se může týkat ověření možnosti konekcionistického systému implementovat určitou kognitivní funkci. Stejně tak by měl mít člověk dopředu jasnou představu o tom, jaký použije experimentální design a jaké analytické nástroje použije ke zhodnocení výsledků experimentu. V tomto velice obecném ohledu se tedy metodologie počítačových simulací příliš neliší od metodologie běžných experimentů prováděných z lidskými subjekty. To, v čem se mezi sebou tyto dva přístupy liší, jsou však určitá specifika jednotlivých dílčích kroků, které musí experimentátor provést v průběhu přípravy experimentální situace. Tyto rozdíly přitom vyplývají zejména z rozdílné povahy „subjektů“, se kterými se pracuje v rámci tradičních experimentů a v rámci počítačových simulací.

2.3.1 Tréninková fáze

Konekcionistické modelování kognitivních procesů má tři hlavní fáze (Plunkett, Elman, 1997): tréninkovou, hodnotící a analytickou. Během tréninkové fáze se snažíme umělou neuronovou síť naučit určitému typu chování či plnit určitou úlohu (*task*). Tou může být například identifikace emocionálního stavu osoby na základě mimiky a výrazu její tváře, převedení psaného textu na plynulý proud řeči, ohodnocení míry rizikovosti nebo bonity bankovního klienta na základě jeho finanční historie atp. Pro umělou neuronovou síť má úloha vždy specifickou podobu souboru vstupně-výstupních dvojic, které tvoří tzv. **tréninkové prostředí** (*training environment*). Umělá neuronová síť (dále UNS) se tak učí pro daný vstup produkovat ten správný výstup. Pro experimentátora z toho vyplývá, že každou úlohu, kterou bude chtít UNS naučit řešit, bude muset nejdříve přeložit do sady dvojic vstupů a odpovídajících výstupů. Vzhledem k tomu, že UNS se učí prostřednictvím indukce na základě prezentovaného souboru konkrétních příkladů cílového chování (rozuměj dvojic vstupů a odpovídajících výstupů), experimentátor musí UNS zajistit přísun odpovídajících informací. Obecně platí zásada, že čím více, tím lépe: UNS jsou nenasytými „žrouty“ informací a čím více informací jim dodáme, tím lepší výsledky nám poskytnou. Důležitá je samozřejmě také kvalita informací, v duchu hesla „*rubbish in, rubbish out*“. V případě učení UNS je důležitá především reprezentativnost poskytnutého souboru příkladů daného chování: Platí celkem triviální zásada, že UNS si dokáže osvojit vždy jen tolik informací, kolik jich je ve vzorku obsaženo, nikdy ne více. V souvislosti s kvalitou dodávaných informací se jako „sympatická“ vlastnost UNS jeví být její schopnost naučit se ignorovat informace, které nejsou důležité či

relevantní z hlediska řešení dané úlohy. I v případě, že UNS dodáme dostatečné množství informací v dostatečné kvalitě, nemáme nikdy zaručeno, že se UNS naučí právě tu generalizaci nebo právě to obecné pravidlo, které chceme, aby by se naučila, neboť z dostupného souboru příkladů lze většinou vyextrahovat větší počet různých generalizací; jistotu nemáme dokonce ani tehdy, když natrénovaná UNS dokáže z každého vstupu tréninkového prostředí vytvořit správný výstup. K ověření toho, co se UNS skutečně naučila slouží druhá a třetí (hodnotící a analytická) fáze simulačního procesu, ve které se s pomocí různých postupů a statistických nástrojů detailně analyzuje chování UNS. Vzhledem k tomu, že nám při konekcionistickém modelování jde především o simulaci kognitivních funkcí lidského mozku, je potřeba brát v úvahu také to, zda jsou informace prezentované UNS v podobě zpětné vazby o chybě ve výstupní vrstvě dostupné rovněž člověku. V případě, že tomu tak není, je nutné to zohlednit při interpretaci výsledného chování UNS jakožto modelu dané kognitivní funkce. S tímto úzce souvisí otázka, které informace experimentátor učiní UNS explicitně dostupné a na které si UNS bude muset přijít takříkajíc sama. Často se konekcionistický model sestavuje právě za tím účelem, aby se ověřila hypotéza, zda UNS dokáže danou informaci z poskytnutých příkladů sama vyextrahovat – v takovém případě daná informace samozřejmě nesmí být v tréninkovém korpusu reprezentována explicitně.

Tímto se dostáváme k další důležité kapitole simulačního procesu, kterou je **způsob reprezentace vstupů a výstupů pro UNS**. Vstupy a výstupy mají pro UNS podobu číselných vektorů, ve kterých každé číslo odpovídá míře aktivace právě jednoho z uzlů ve vstupní, resp. výstupní vrstvě UNS. V závislosti na použité aktivační funkci se jednotlivé číselné hodnoty mohou pohybovat v různém rozmezí, většinou však mezi 0 a 1. Reprezentace nějakého vstupu takto může mít například podobu vektoru $[1 \ 1 \ 0 \ 1 \ 0]$. Tímto způsobem pak musí experimentátor přereprezentovat UNS každý vstup a každý výstup z jejího tréninkového prostředí. Právě v této fázi činí experimentátor rozhodnutí (nebo by si alespoň měl být vědom, že ho činí) o tom, jaké informace dá UNS explicitně k dispozici. Například slova, která budou sloužit jako vstup do nějaké úlohy, mohou být reprezentována vektorem, který se skládá ze souboru 0 a 1, které budou ve zcela náhodném pořadí, takže z nich UNS nedokáže vyčíst žádnou užitečnou informaci, kterou by mohla využít k řešení dané úlohy. Při prezentaci určitého slova-vektoru na vstupní vrstvě tak UNS bude vědět pouze to, že takové slovo-vektor existuje. Experimentátor se však může rozhodnout, že slova bude reprezentovat jako soubor jednotlivých hlásek, z nichž každé bude odpovídat ve vektoru jedno číslo, 0 nebo 1, podle toho, zda je daná hláska v daném

slově přítomná (1) či nikoli (0).¹ V případě použití takového způsobu reprezentace by UNS disponovala řadou dalších informací, z nichž některé by se mohly v průběhu učení ukázat jako užitečné při řešení dané úlohy (nebo naopak jako zcela irelevantní a nedůležité). Jednotlivá slova by také bylo možné reprezentovat souborem vektorů, z nichž by každý reprezentoval jednotlivé hlásky jako soubor mnoha dílčích vlastností typu *znělosti*, *drsnosti*, *kompaktnosti* atp. Tento způsob reprezentace by dal UNS k dispozici zase trochu jiný soubor informací, které by se UNS mohla – díky své schopnosti integrovat a zohlednit najednou velké množství dílčích informací, vodítek a omezení – pokusit využít při řešení dané úlohy.

V souvislosti s volbou způsobu reprezentace vstupů a výstupů pro UNS je zde důležité zmínit, že toto kódování experimentátor většinou provádí v rámci tzv. lokálního nebo rysového reprezentačního schématu (*local/featural representational scheme*) a to čistě z důvodu uživatelské vstřícnosti rozhraní mezi experimentátorem a UNS. Jak již bylo uvedeno v jednom z oddílů věnovaných charakteristikám PDP modelů, pro tento typ výpočetní architektury jsou charakteristické distribuované reprezentace, jejichž základní jednotky (jednotlivé uzly sítě) nekorespondují s žádným z objektů, o kterých se hovoří v rámci běžného mentalistického diskurzu. To jinými slovy znamená, že pro člověka je tento typ reprezentace sémanticky zcela neprůhledný. Proto se ke kódování vstupů a výstupů pro UNS používají dva odlišné typy reprezentací, které jsou již pro člověka mnohem srozumitelnější.

Jedním typem jsou tzv. **lokální reprezentace** (*local representations*), kdy jednotlivé uzly sítě odpovídají běžným pojmům nebo myšlenkám. Takto může například aktivační vektor [0 1 0 0] reprezentovat pojem psa. Při použití tohoto způsobu kódování stačí k identifikaci reprezentovaného pojmu lokalizovat pozici jediného aktivního uzlu. Druhým typem jsou tzv. **rysové reprezentace** (*featural representations*), které jsou druhem distribuované reprezentace, kdy jednotlivé uzly sítě kódují různé dílčí vlastnosti, které dohromady reprezentují nějaký běžný pojem. Například pojem psa může být reprezentován aktivačním vektorem [1 1 1 0 0 1], kde míra aktivace sedmi uzlů sítě kóduje vlastnosti typu „chlupatý“, „čtyřnohý“, „štěkající“, „dvounohý“, „mňoukající“, „věrný“ a „nebezpečný“. K identifikaci reprezentovaného pojmu psa zde nestačí znát míru aktivace pouze jednoho uzlu, ale je potřeba znát míru aktivace všech sedmi uzlů, které se na reprezentaci tohoto pojmu podílejí. Tento typ distribuované reprezentace je na rozdíl od té, kterou můžeme nalézt například ve skryté vrstvě UNS, sémanticky zcela transparentní, tzn. že jednotlivým uzlům lze přiřadit významy, se kterými se běžně operuje v každodenním diskurzu (Lloyd, 1989; Plunkett, Elman, 1997). Alespoň určitý vhled do

¹ Kódování může být samozřejmě i zcela opačné, důležité je pouze systematické používání zvoleného kódování.

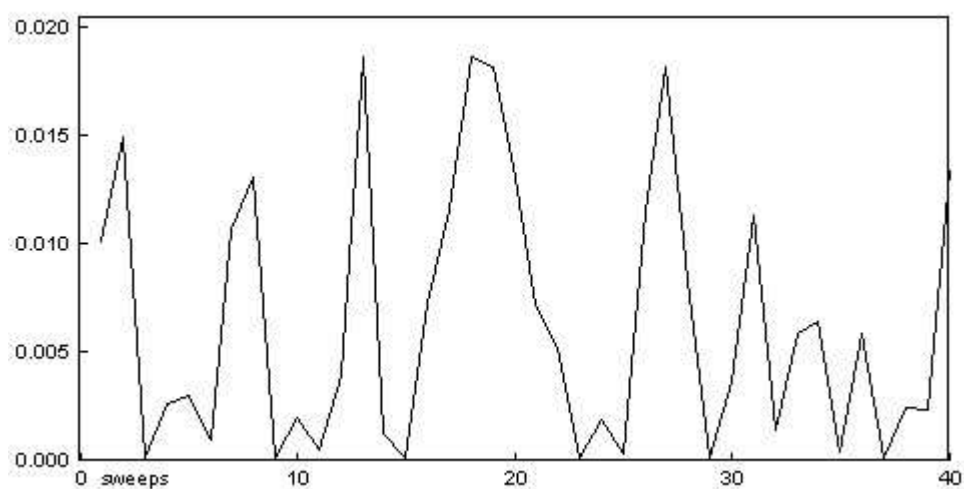
sémanticky neprůhledných reprezentací ve skryté vrstvě neuronové sítě nabízí různé statistické nástroje multivariační analýzy dat, které jsou podrobněji zmíněny v oddíle věnovaném analytické části simulačního experimentu.

Dalším důležitým úkolem experimentátora je **volba správné architektury UNS**, která by se dokázala naučit řešit požadovanou úlohu. Architektura UNS je dána počtem uzlů sítě a způsobem jejich zapojení. Počet uzlů ve vstupní a výstupní vrstvě je determinován rozhodnutími učiněnými v souvislosti se způsobem reprezentace vstupně-výstupních dvojic tréninkového prostředí – použijí-li k reprezentaci vstupu vektor o délce 6, potom budu na vstupní vrstvě UNS potřebovat právě šest uzlů. V případě skryté vrstvy je výběr počtu uzlů plavbou mezi *Skyllou* a *Charybdou*, neboť na jedné straně příliš mnoho uzlů vede k přeučení (*overlearning* nebo také *overfitting*), kdy se UNS naučí vstupně-výstupní dvojice tréninkového prostředí mechanicky nazpaměť, ale neosvojí si obecnější pravidlo, které by jí umožnilo generalizovat i na dosud nepotkané případy, a na straně druhé příliš málo uzlů vede k neschopnosti generalizace z důvodu nedostatku výpočetních zdrojů nutných k extrakci a zakódování obecného pravidla. Způsob zapojení uzlů sítě (například do jednoduché dopředné, rekurentní či atraktorové sítě), pak vychází z experimentátorových předpokladů ohledně způsobů zpracování informací, které jsou podle něj nutné k vyřešení dané úlohy. Obecně se dá říci, že volba správné architektury sítě se řídí metodou pokusu a omylu, kterou je však časem možné doplnit praxí získanými zkušenostmi a intuicí.

2.3.2 Hodnotící fáze

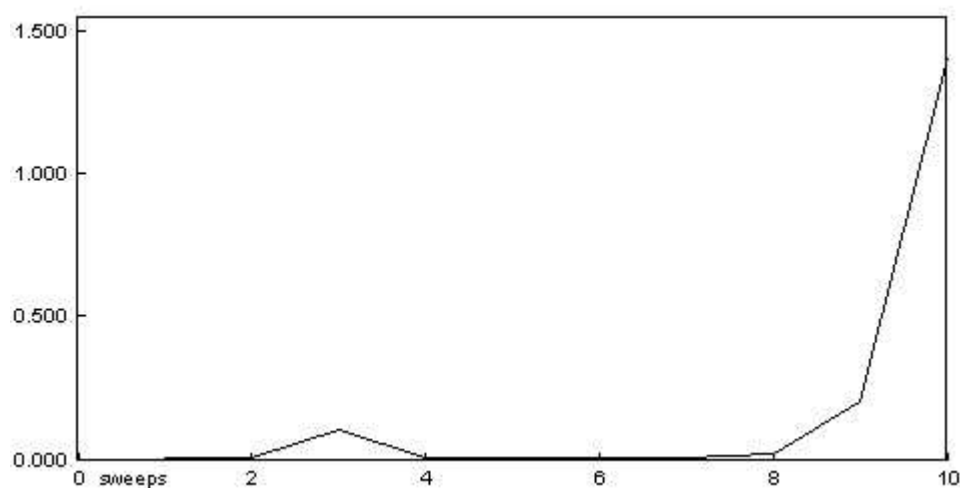
Po tréninkové fázi následuje fáze hodnotící, jejíž náplní je hodnocení výkonu natrénované UNS. K hodnocení výkonu UNS se využívají především dva ukazatele: **globální chyba** (*global error*) a **vzorec individuálních chyb** (*individual pattern error*). V průběhu trénování UNS simulátor počítá chybu na jejím výstupu podle jednoduchého vzorce, který od požadovaného výstupu (v podobě nějakého výstupního vektoru) odečítá skutečný výstup, který UNS vygenerovala na základě prezentovaného vstupu. Při odečítání vektorů se postupuje jednoduše tak, že se vždy od sebe odečtou ta čísla ze dvou vektorů, která se v nich nacházejí na stejné relativní pozici, tzn. že první se odečte od prvního, druhý od druhého atd. Vzhledem k tomu, že hodnota rozdílu mezi jednotlivými čísly vektoru může být pozitivní, nebo negativní, pracuje se s absolutními hodnotami nebo s odmocninami druhých mocnin zjištěných rozdílů. Chyba na daném výstupu je pak suma absolutních hodnot (nebo odmocnin čtverce) těchto rozdílů. Takto například může mít požadovaný a skutečný výstupní vektor podobu [1 1 0 1 0], resp.

[0,5 0 1 1 0]; v takovém případě chyba na výstupu bude mít hodnotu 2,5, neboť součet absolutních hodnot rozdílů všech pěti čísel ve vektoru (0,5, 1, 1, 0, 0) je roven právě 2,5. Tímto způsobem lze zjistit, jak si UNS vede při transformaci jednotlivých vstupních vektorů na výstupní vektory. Zjištěné chyby se mohou týkat vstupně-výstupních dvojic, které jsou obsaženy v tréninkovém prostředí, nebo dvojic vstupů a výstupů, které se používají k testování schopnosti UNS generalizovat určitým konkrétním způsobem na nové a dosud nepotkané případy. Výsledný graf má podobu **vzorce individuálních chyb**, jehož příklad je možné vidět v grafech 15 a 16, na kterých je zachycen vzorec individuálních chyb UNS, která byla trénována za účelem rozlišování a predikce dobrých a špatných zaměstnanců na základě jejich psychologických profilů zjištěných pomocí Hoganova osobnostního dotazníku.¹ Vzhledem k tomu, že konekcionistické výpočetní modely se díky své schopnosti se učit často používají



Graf 15: Vzorec individuálních chyb umělé neuronové sítě, která byla trénována za účelem rozlišování dobrých a špatných zaměstnanců na základě jejich psychologických profilů zjištěných pomocí Hoganova osobnostního dotazníku. V grafu je zachycena úspěšnost, s jakou UNS dokáže správně zařadit 40 zaměstnanců, kteří sloužili jako tréninkové prostředí. UNS se v psychologických profilech dvaceti dobrých a dvaceti špatných zaměstnanců snažila najít zákonitosti, které by bylo možné využít k predikci výkonu budoucích zaměstnanců-uchazečů. Na grafu je dobře vidět, že chyby na výstupu jsou minimální ($< 0,02$), tzn. že UNS se podařilo využít struktury obsažené v psychologických profilech k rozřazení zaměstnanců do dvou výkonnostních kategorií. Je ovšem otázka, zda se UNS podařilo objevit nějaké obecnější pravidlo, které by bylo možné úspěšně aplikovat i na další, dosud nepotkané případy, nebo zda se jen mechanicky naučila nazpaměť, který ze 40 zaměstnanců patří do kategorie „dobrý zaměstnanec“ a který do kategorie „špatný zaměstnanec“. To lze ověřit tím, že se natrénovaná UNS použije na nové případy, u kterých přitom víme, jak by měl vypadat správný výstup...

¹ Tento predikční model na bázi umělé neuronové sítě jsem vytvořil v rámci své praxe ve společnosti *Assessment Systems*, s. r. o., která mi poskytla veškerá data potřebná k jeho sestavení.



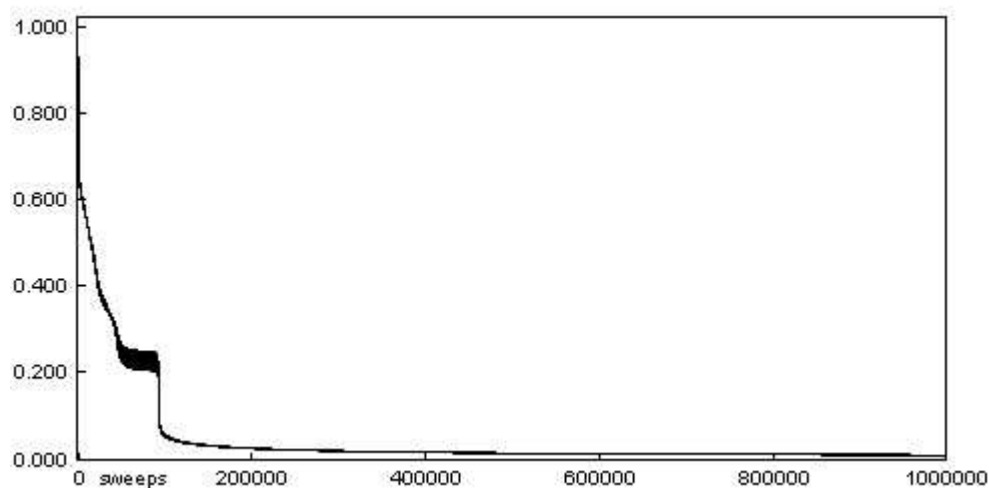
Graf 16: ...a zde nám vzorec individuálních chyb ukazuje, že natrénovaná UNS dokáže správně zařadit 9 z 10 dosud nepotkaných případů-zaměstnanců. To naznačuje, že UNS se podařilo osvojit si určité obecnější pravidlo, které jí umožňuje se správně rozhodovat i v nových situacích, se kterými nemá žádnou „osobní“ zkušenost. (oba grafy byly vytvořeny s pomocí simulačního softwaru Tlearn.)

jako modely procesu osvojování různých kognitivních funkcí, nachází chybový ukazatel vzorce individuálních chyb hojného využití zejména při porovnávání výkonu UNS v jednotlivých fázích jejího tréninku, kdy UNS disponuje různými zkušenosti a v důsledku toho tedy také různými váhovými maticemi determinujícími její výkon.

K hodnocení výkonu UNS se vedle vzorce individuálních chyb používá také tzv. **globální chyba**, která je celkovým ukazatelem chybovosti UNS napříč celým souborem vstupně-výstupních dvojic. Globální chyba se často počítá v podobě druhé odmocniny průměru čtverce všech dílčích výstupních chyb (tzv. *RMS*, *Root Mean Square error*), tj. podle vzorce

$$rms = \sqrt{\frac{\sum_k \left(\vec{t}_k - \vec{o}_k \right)^2}{k}},$$

kde k je počet vstupních vektorů (vzorců aktivace), \vec{o}_k je výstupní vektor vygenerovaný UNS na základě prezentovaného vstupního vektoru a \vec{t}_k je požadovaný (cílový) výstupní vektor. Příklad grafického vyjádření globální chyby je graf 17, kde je dobře vidět, jak celková míra chybovosti UNS v průběhu učení postupně klesá.



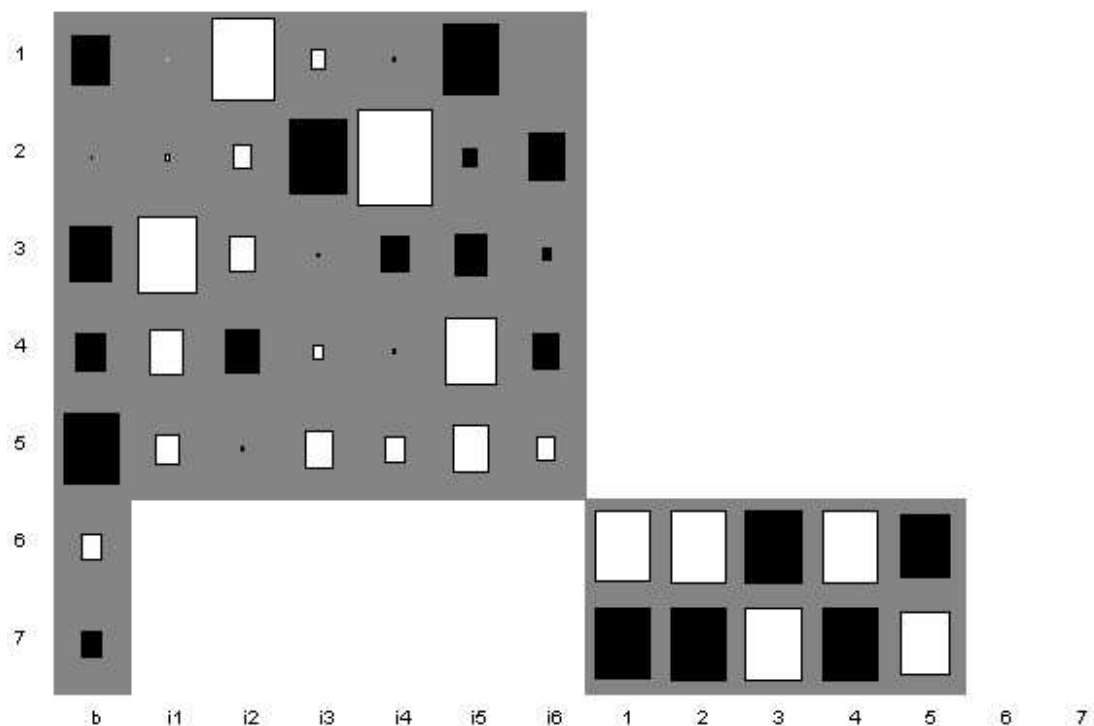
Graf 17: Globální chyba umělé neuronové sítě. Z grafu je dobře patrné, jak se celková míra chybovosti UNS v průběhu učení postupně snižuje. Nejvyšší míra chybovosti je na samotném začátku, kdy UNS začíná s náhodně nastavenými vahami jednotlivých spojů. S jejich postupnými úpravami se výkon UNS zlepšuje, nejdříve rychle, ve skocích, poté již mnohem pomaleji a jen velice pozvolna. Každá ze 40 vstupně-výstupních dvojic tréninkového korpusu byla UNS prezentována 25 000 x (1 000 000/40); tuto skutečnost by také bylo možné vyjádřit tak, že trénování UNS probíhalo v 25 000 epochách, které představují „jeden průběh [celé] množiny dat v iterativním procesu trénování neuronové sítě“ (Fanta, 2000, s. 26). (vytvořeno s pomocí simulačního softwaru Tlearn)

2.3.1 Analytická fáze

Ukazatele chybovosti, zejména pak vzorec individuálních chyb lze použít i ve třetí fázi simulačního procesu, ve které se analyzuje, co se UNS ve skutečnosti naučila. Jak zde již bylo uvedeno, UNS se z prezentovaných příkladů může ve většině případů naučit několik různých generalizací či pravidel a nikdy není zaručeno, že se naučí právě tu generalizaci, kterou chceme, aby se naučila. K tomu, aby se zjistilo, kterou konkrétní generalizaci si UNS osvojila, se často používá metoda, která spočívá v systematickém vystavování UNS souboru nových podnětů a ve sledování, jak UNS na tyto podněty reaguje. A právě k porovnávání reakcí UNS na jednotlivé vstupní podněty je ideálním nástrojem vzorec individuálních chyb. Tento postup je de facto identický s tím, který se běžně využívá v rámci kognitivně-psychologického výzkumu, kdy se u lidských subjektů prostřednictvím prezentace podnětového materiálu vyvolává určité chování a na jeho základě se pak usuzuje na charakteristiky vnitřních kognitivních procesů, které toto chování produkují. Na rozdíl od kognitivní psychologie má však metodologie konekcionistického modelování kognitivních

procesů k dispozici řadu dalších nástrojů, které jí umožňují snadno nahlédnout přímo do samotných „vnitřností“ mechanismu, který produkuje pozorovatelné chování UNS.

Vzhledem k tomu, že zpracovávání informací (tj. jejich manipulace a transformace) je v UNS zprostředkováno především váženými spoji mezi jejími jednotlivými uzly, je dobré mít při analýze chování UNS k dispozici nějaký nástroj, který by umožňoval přehlednou vizualizaci jinak velice spleťitého přediva vážených spojů. Tuto funkci plní tzv. **Hintonův diagram**, ve kterém je relativní síla váženého spoje reprezentována velikostí obdélníka, jeho „náboj“ barvou obdélníka (negativní/inhibiční – černá, pozitivní/excitační – bílá) a jeho orientace prostřednictvím řádků (označujících přijímající uzel) a sloupců (označujících vysílající uzel) (viz obrázek 151).



Obrázek 151: Hintonův diagram znázorňující relativně přehledným způsobem váhovou matici umělé neuronové sítě. (vytvořeno s pomocí simulačního softwaru Tlearn)

To, jak UNS „rozumí“ prezentovaným podnětům a jak je zpracovává, se neprojevuje pouze ve váhové matici kódující její „mentální“ procesy, ale také ve způsobu, jakým si UNS re-reprezentuje vstupní data ve své skryté vrstvě, jejímž prostřednictvím mění strukturu podobnosti vstupních vzorců aktivace za účelem požadované transformace vstupů na výstupy. Velká pozornost je proto při analýze chování UNS věnována také jejím vnitřním „mentálním“ reprezentacím: UNS se prezentací vstupních vzorců přiměje k tomu, aby ve své skryté vrstvě

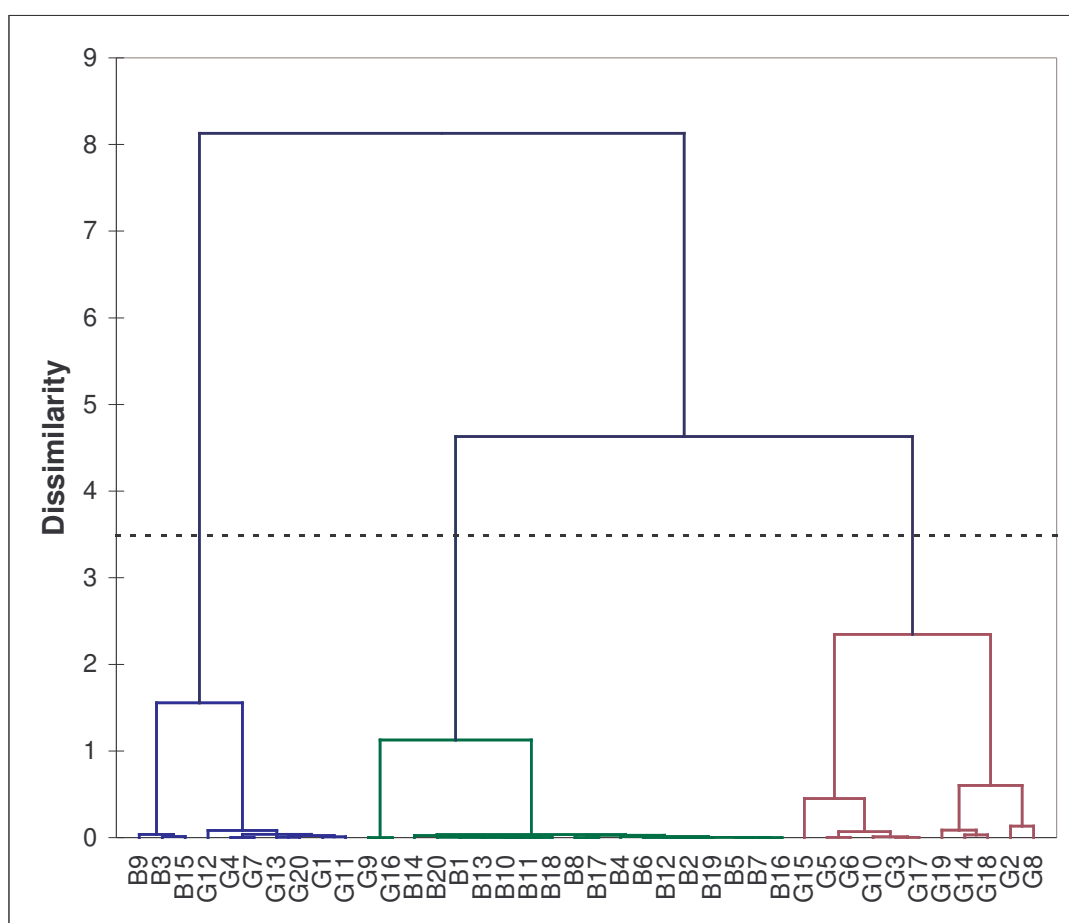
aktivovala jejich vnitřní reprezentace, které tedy nejsou ničím jiným než vektorci (vektory) aktivace ve skryté vrstvě. Tyto vzorce aktivace definují n -rozměrný prostor, kde n je počet uzlů ve skryté vrstvě. Aktivační vektory ve skryté vrstvě vyvolané jednotlivými vstupními podněty pak specifikují přesné souřadnice těchto vstupních podnětů v n -rozměrném prostoru. Způsob, jakým jsou jednotlivé vstupní podněty v tomto n -rozměrném prostoru rozmístěny, tj. jak blízko sebe se jednotlivé podněty nacházejí, pak vypovídá o tom, jak UNS těmto podnětům rozumí – ve smyslu jejich ne/podobnosti z hlediska určitého kritéria, které se jí „zdá“ být podstatné pro řešení dané úlohy (transformace vstupů na výstupy); přitom platí, že čím jsou si dané vstupní podněty (podle UNS) podobnější, tím k sobě mají v n -rozměrném prostoru blíže (ve smyslu obvyklé euklidovské vzdálenosti). Analýza struktury podobnosti tohoto vnitřního prostoru takto může odhalit, proč se UNS chová tak, jak se chová. Vzhledem k tomu, že ve skryté vrstvě UNS se většinou vyskytuje větší množství uzlů (více než tři), představuje přímá vizualizace vnitřního reprezentačního prostoru, jehož počet dimenzí odpovídá počtu uzlů ve skryté vrstvě, značný problém, který se však daří - více či méně úspěšně - obcházet s pomocí řady různých statistických nástrojů jako je **hierarchická shluková analýza, multidimenzionální škálování** či **analýza hlavních komponent**.

První dvě metody vycházejí z matice podobností (nebo rozdílů) mezi jednotlivými aktivačními vektory (jakési obdoby tabulek vzdáleností mezi městy v autoatlase). Míru podobnosti jednotlivých vektorů aktivace mezi sebou je možné vyjádřit například prostřednictvím tzv. *D-koefficientu* (Ferjenčík, 2000, s. 194), který se počítá podle vzorce

$$D_{ab} = \sqrt{\sum_{i=1}^n (d_{ab})^2},$$

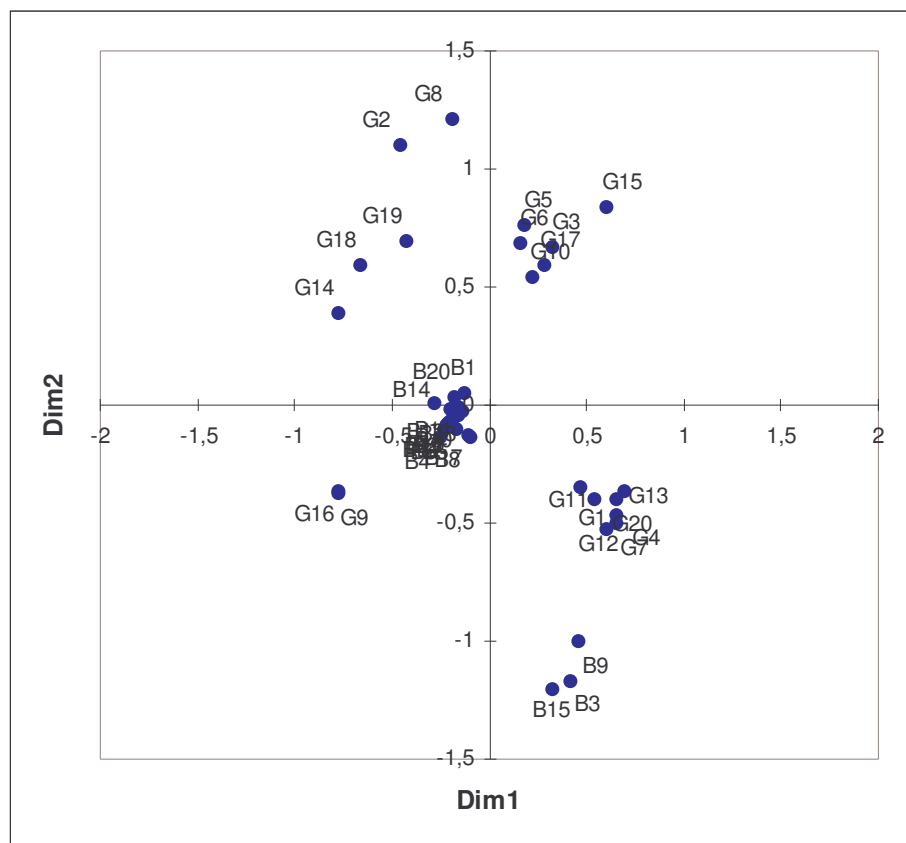
kde D_{ab} je koeficient celkové podobnosti mezi vektory A a B, n je počet uzlů ve skryté vrstvě a d_{ab} je rozdíl mezi čísly ze dvou vektorů, která se v nich nacházejí na stejné relativní pozici; platí přitom, že čím vyšších hodnot D-koefficient nabývá, tím menší je podobnost mezi dvěma porovnávanými aktivačními vektory. Výsledná matice podobností je pak vstupem do výpočetního algoritmu hierarchické shlukové analýzy (HSA) či multidimenzionálního škálování (MDS), který za nás provede počítač. Rozdíl mezi oběma statistickými nástroji spočívá zejména ve způsobu vizualizace výsledné struktury ne/podobnosti vnitřního reprezentačního prostoru: Zatímco výstupem HSA je dendrogram, ve kterém se nepodobné aktivační vektory nacházejí na vzdálenějších „větších“ a podobné aktivační vzorce na bližších a níže položených „větších“ (viz graf 18), výstupem MS je matice dimenzí, kterou lze (v

případě, že výstupem jsou dvě nebo tři dimenze) snadno převést do podoby běžného 2-D nebo 3-D grafu, který člověku umožňuje přímo si prohlížet některé (z hlediska řešené úlohy) zajímavé řezy vícedimenzionálním prostorem vnitřních reprezentací UNS (viz graf 19 a 20). Při použití MDS je dobré si být vědom toho, že je s ním vždy spojen jistý kompromis mezi uživatelskou vstřícností a přesností: S pomocí MDS lze původní počet dimenzí zredukovat v libovolné míře, platí však pravidlo, že čím více dimenzí, tím větší míra shody mezi novým zobrazením a skutečnými vzdálenostmi, avšak tím obtížnější interpretace, a čím méně dimenzí (dvě nebo tři), tím snazší interpretace, ale tím vyšší riziko výrazného zkreslení reálných vzdáleností.¹

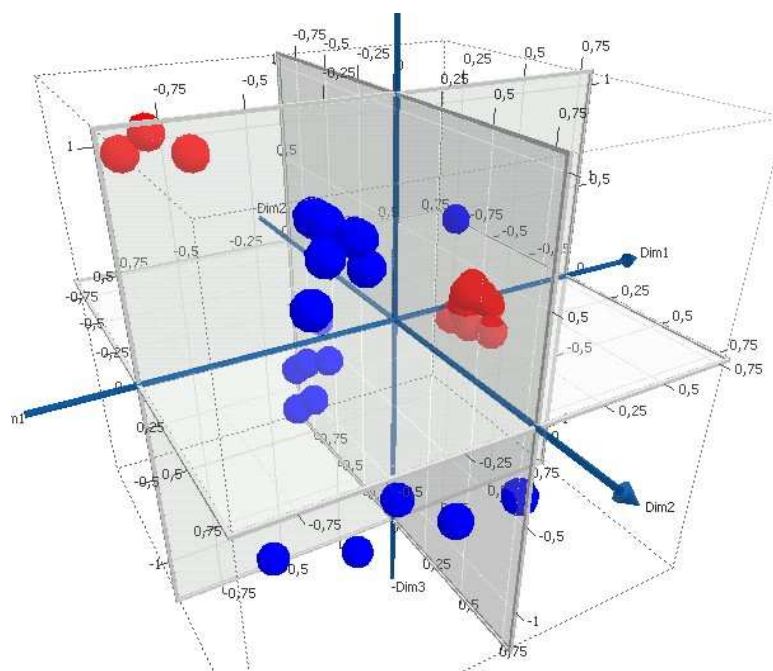


Graf 18: Dendrogram hierarchické shlukové analýzy vnitřního reprezentačního prostoru umělé neuronové sítě trénované na rozlišování a predikci dobrých (G) a špatných (B) zaměstnanců na základě jejich psychologických profilů zjišťovaných s pomocí Hoganova osobnostního dotazníku.

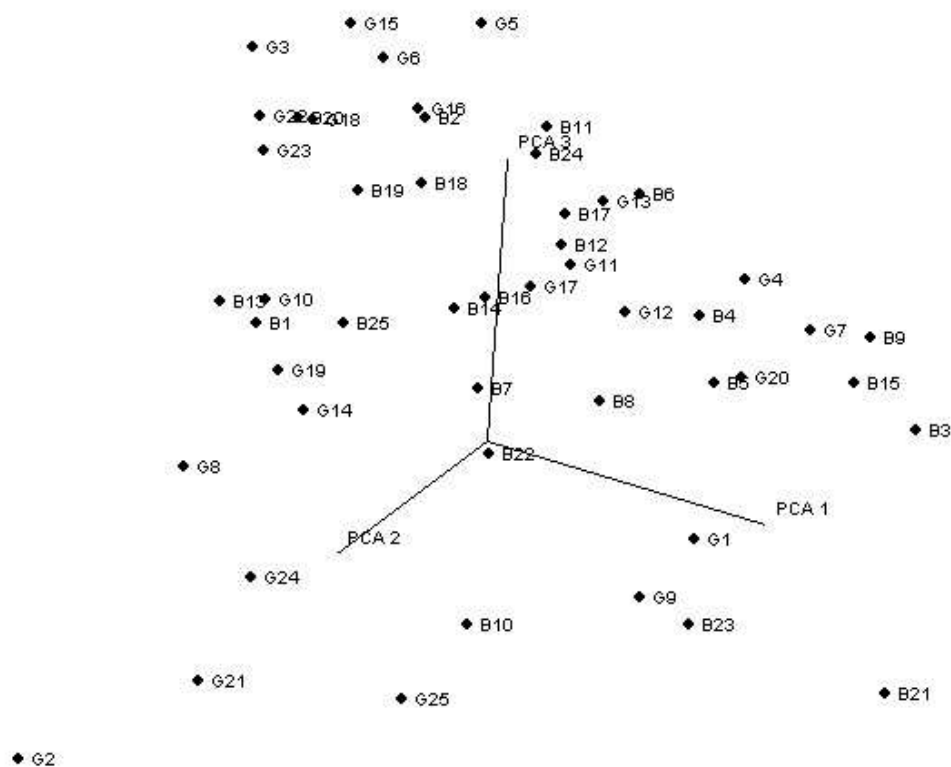
¹ To, o co se MDS snaží, je nalézt takové rozložení bodů v prostoru s menším počtem dimenzí, které by co možná nejvíce zachovalo relativní euklidovské vzdálenosti naměřené mezi body ve vícedimenzionálním prostoru. Při tomto “smrštění” logicky dochází k určitému zkreslení těchto relativních vzdáleností. Míru tohoto nesouladu mezi původním a novým zobrazením vyjadřuje tzv. stresová funkce; platí přitom, že dokud hodnota této stresové funkce nepřekročí 0,2, můžeme hovořit o ještě přijatelné míře zkreslení (Kruskal, Wish, 1978).



Graf 19: 2-D graf vnitřního reprezentačního prostoru umělé neuronové sítě vytvořený na základě multidimenzionálního škálování (G- dobří zaměstnanci, B – špatní zaměstnanci). Z grafu je dobře patrné, jak si UNS zaměstnance rozdělila do několika jasně rozlišitelných skupinek dobrých a špatných zaměstnanců.



Graf 20: 3-D graf vnitřního reprezentačního prostoru umělé neuronové sítě vytvořený na základě multidimenzionálního škálování (modrá barva- dobří zaměstnanci, červená barva – špatní zaměstnanci).

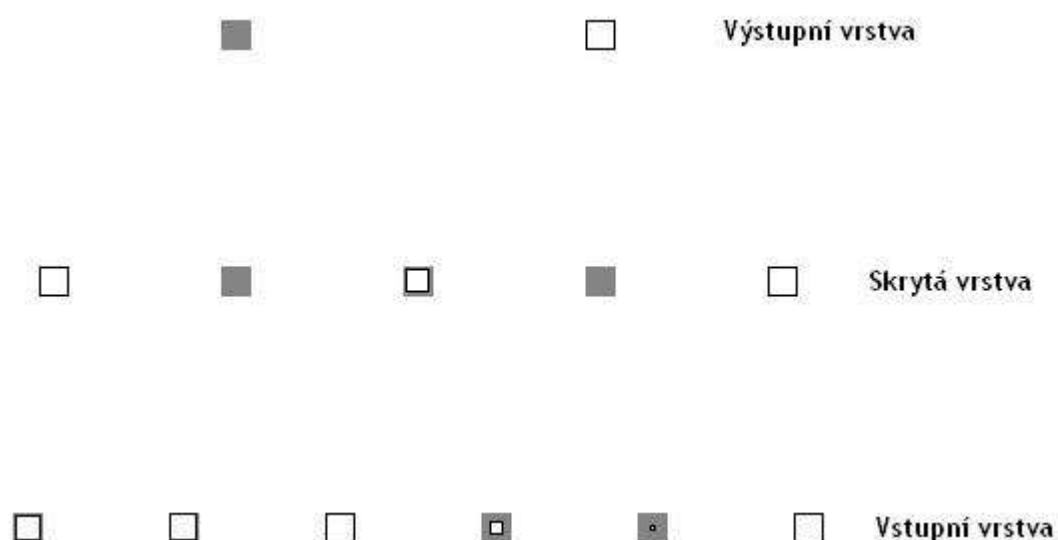


Graf 21: 3-D graf vnitřního reprezentačního prostoru umělé neuronové sítě vytvořený na základě analýzy hlavních komponent (G – dobří, B – špatní zaměstnanci). (vytvořeno s pomocí simulačního softwaru Tlearn)

Další v tomto kontextu často používanou metodou je analýza hlavních komponent (AHK). Tato metoda vychází z korelační matice - obdobu matice podobností, která se liší jen tím, že vztahy mezi aktivačními vektory v ní nejsou zaneseny v podobě koeficientů podobnosti, ale v podobě korelačních koeficientů. Výstupem AHK jsou faktory, které vysvětlují variabilitu většího počtu proměnných jejich uspořádáním do několika málo větších uskupení. AHK tak podobně jako faktorová analýza umožňuje vysvětlit velké množství viditelných proměnných menším počtem základních proměnných (faktorů), které nejsou přímo pozorovatelné. V případě použití dvou nebo tří faktorů lze výsledek přehledným způsobem zobrazit v podobě 2-D nebo 3-D grafu (viz graf 21). Pro AHK platí obdobné omezení jako pro MDS: Čím větší počet faktorů, tím více variability proměnných lze vysvětlit, avšak tím obtížnější je se v nepřehledném množství proměnných zorientovat, a čím méně faktorů, tím snazší orientace, ale tím více variability proměnných zůstane nevysvětleno.

Vedle těchto způsobů vizualizace vnitřního reprezentačního prostoru UNS, které využívají různých sofistikovaných statistických nástrojů, se při analýze vnitřních reprezentací UNS využívá rovněž jednoduché grafické znázornění míry aktivace uzlů v jednotlivých vrstvách UNS (viz obrázek 152). To experimentátorovi umožňuje zkoumat, jak UNS ve své skryté a

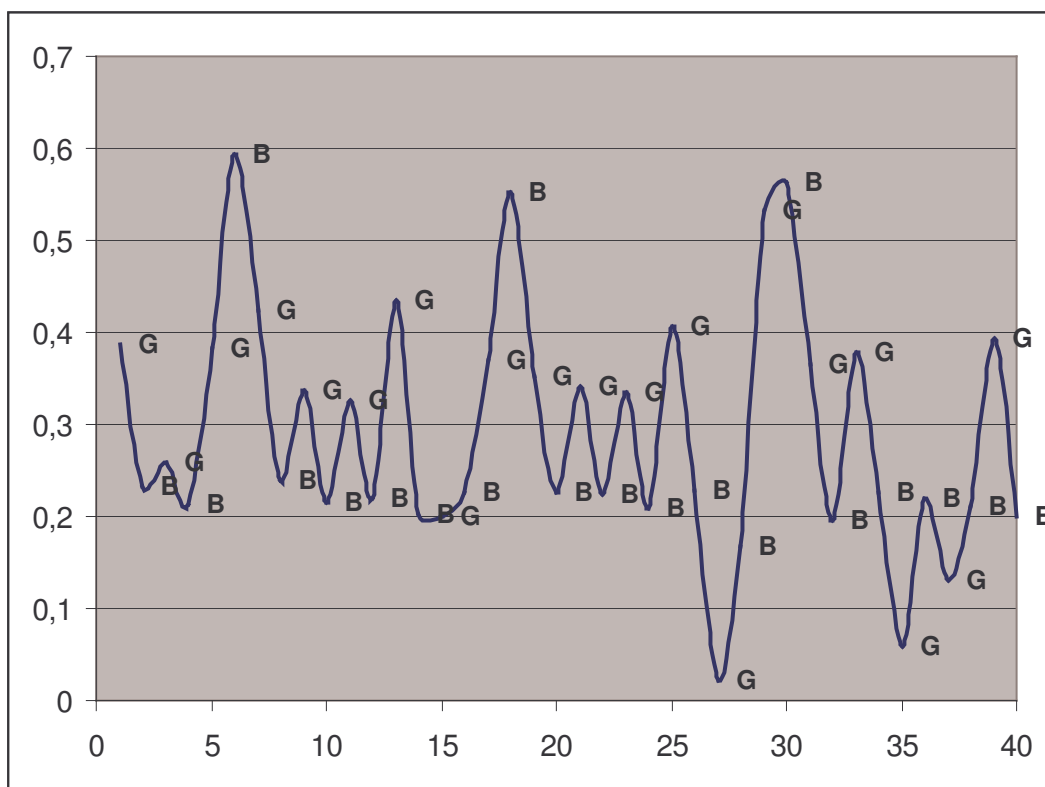
výstupní vrstvě reaguje na různé vstupní podněty. Jedná se de facto o jakousi obdobu funkčních zobrazovacích metod, které se v rámci kognitivní neurovědy využívají ke zjišťování, jaké oblasti mozku se aktivují nebo naopak tlumí svou činnost při řešení různých druhů úloh.



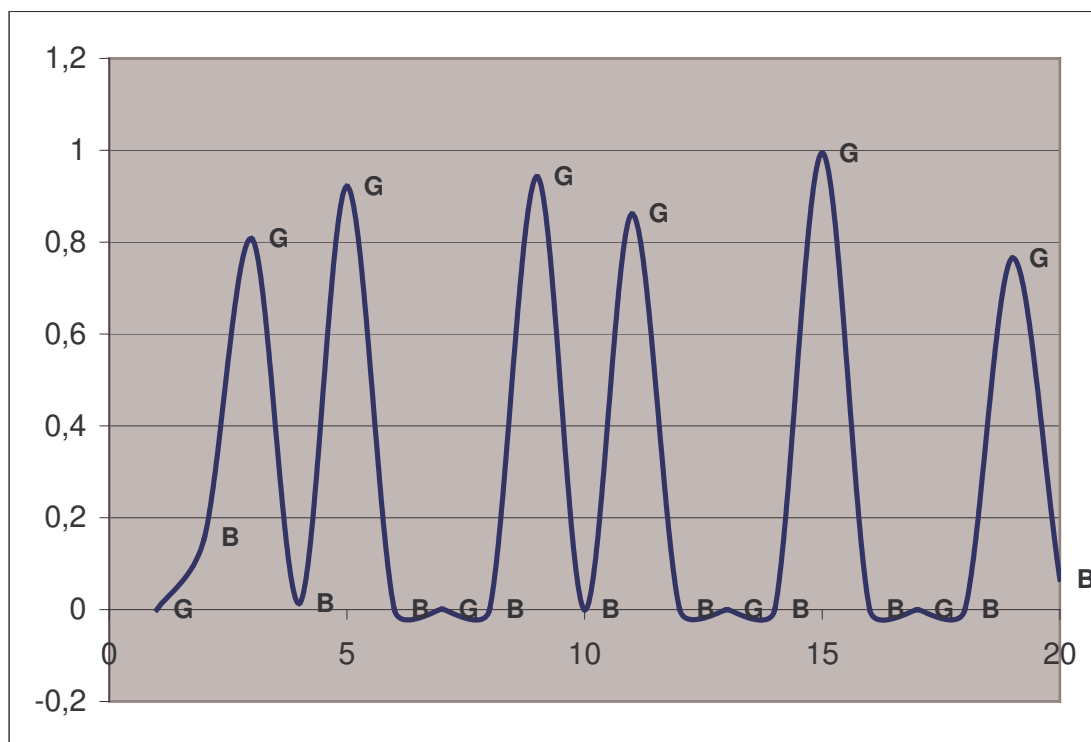
Obrázek 152: Zobrazení míry aktivace uzlů v jednotlivých vrstvách umělé neuronové sítě. Míra aktivace uzlu je zde reprezentována velikostí bílého čtverce: čím více bílé barvy, tím vyšší míra aktivace a čím více šedivé barvy, tím nižší míra aktivace. Zobrazení funguje jako ekvivalent funkčních zobrazovacích metod používaných v rámci kognitivní neurovědy ke zjišťování míry námahy jednotlivých oblastí mozku při řešení různých druhů úloh. (vytvořeno s pomocí simulačního softwaru Tlearn)

Při analýze chování UNS lze použít také ekvivalenty dalších metod a nástrojů, které se jinak běžně používají při zkoumání biologického mozku (Lloyd, 2004): Takto lze například graficky zachytit vývoj průměrné míry aktivace všech uzlů ve skryté vrstvě UNS v závislosti na druhu prezentovaného podnětu. Výsledný grafický záznam odpovídá záznamu klasického elektroencefalografického (EEG) vyšetření metodou **evokovaných potenciálů** (*evoked* či *event-related potentials*), kdy se s pomocí elektrod připevněných na vlasatou část hlavy snímá elektrická aktivita rozsáhlé populace nervových buněk (nacházejících se pod danou elektrodou) během prezentace různých druhů podnětů (viz graf 22). Ekvivalentem **mikroelektrod** zavedených přímo do mozku a snímajících elektrickou aktivitu jednotlivých neuronů je pak graf zaznamenávající vývoj míry aktivace jednotlivých uzlů ve skryté vrstvě UNS (viz graf 23 a 24). K zobrazení dynamiky chování UNS (její vývojové trajektorie) lze také použít 2-D či 3-D stavový diagram, jehož jednotlivé dimenze odpovídají buď míře

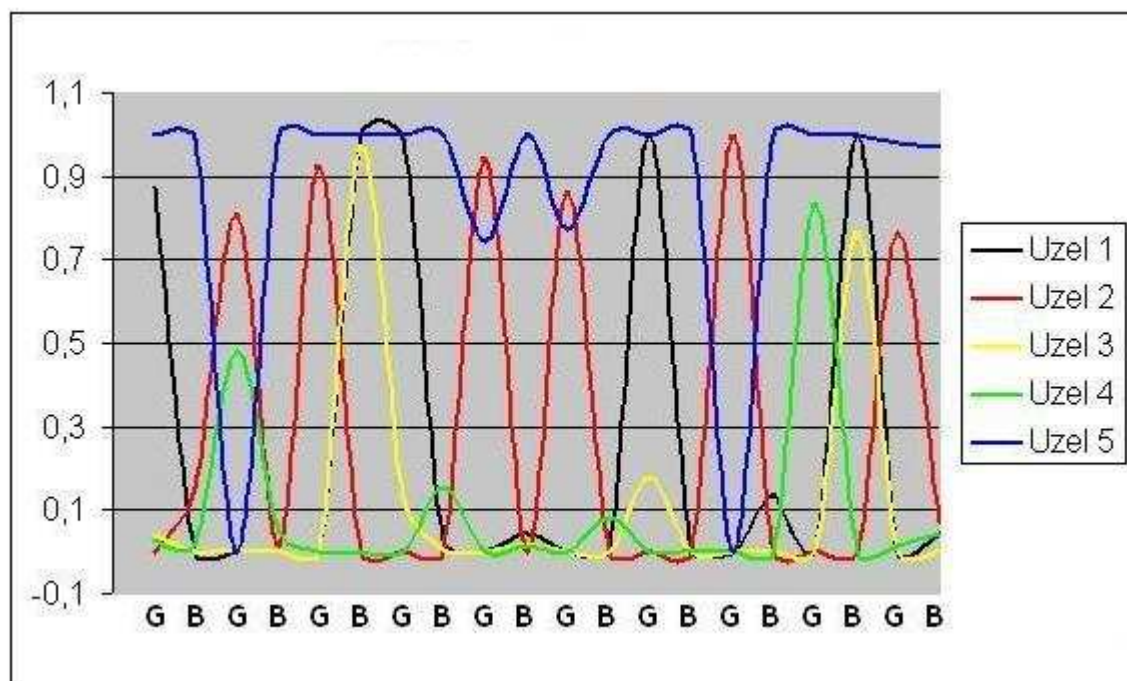
aktivity vybraných uzlů skryté vrstvy během prezentace jednotlivých podnětů, nebo výstupům vybraného typu multivariačního zpracování aktivačních vektorů skryté vrstvy, která obsahuje více než tři uzly (viz graf 25).



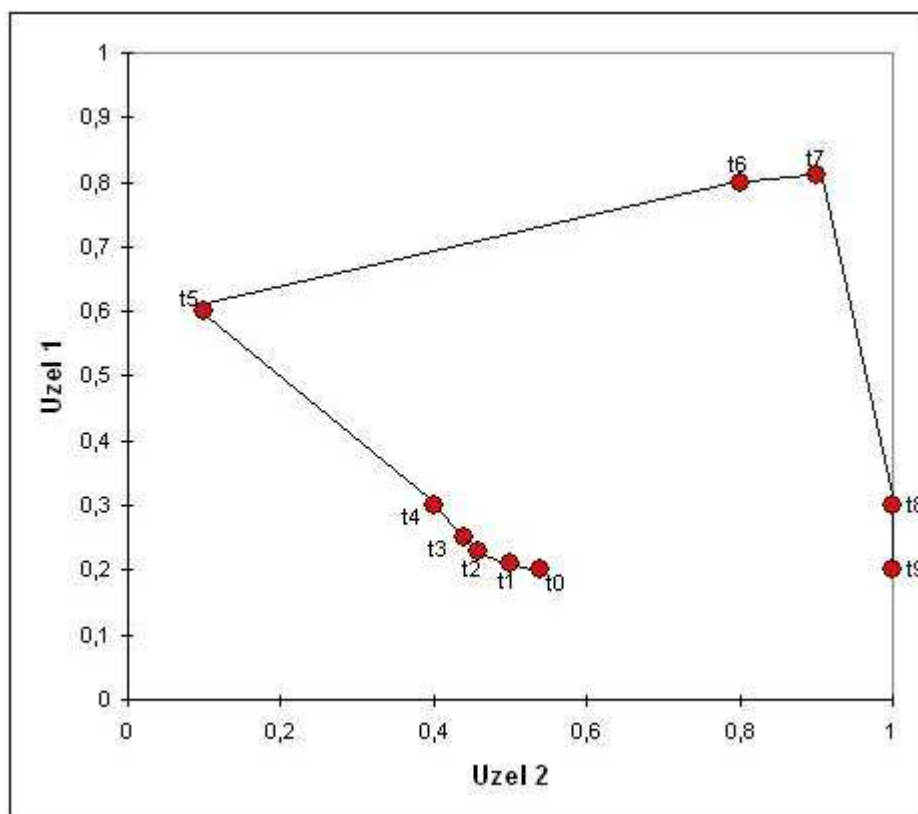
Graf 22: Průběh „EEG“ křivky umělé neuronové sítě v závislosti na prezentovaném podnětu („evokované potenciály“). Jedná se o grafický záznam průměrné míry aktivace pěti uzlů ve skryté vrstvě UNS zpracovávající psychologické profily zaměstnanců za účelem predikce jejich budoucího výkonu. Z grafu jsou celkem dobře patrné dvě základní tendence ve vývoji „EEG“ křivky: 1) Při zpracovávání a rozpoznávání psychologického profilu většiny dobrých zaměstnanců (G) je průměrná míra aktivace skryté vrstvy relativně vyšší než v případě zpracovávání psychologického profilu většiny špatných zaměstnanců (B). 2) V několika případech je tomu však přesně naopak: Výrazně vysoká (průměrná) míra aktivace uzlů ve skryté vrstvě je spojena se zpracováváním psychologického profilu špatných zaměstnanců a výrazně nízká míra aktivace je zase spojena s rozpoznáváním psychologického profilu dobrých zaměstnanců.



Graf 23: Grafický záznam průběhu míry aktivace jednoho uzlu ve skryté vrstvě v závislosti na prezentovaném podnětu. Z grafu je dobře patrné, že při zpracovávání psychologického profilu většiny dobrých zaměstnanců (G) má tento konkrétní uzel tendenci zvyšovat míru své aktivity a že při zpracovávání psychologického profilu špatných zaměstnanců (B) má naopak tendenci svou činnost tlumit.



Graf 24: Grafický záznam průběhu míry aktivace všech pěti uzlů ve skryté vrstvě v závislosti na prezentovaném podnětu (dobrý/špatný zaměstnanec, G/B).



Graf 25: Stavový diagram zachycující vývojovou trajektorii UNS definovanou mírou aktivace dvou uzlů skryté vrstvy v závislosti na prezentovaných podnětech v čase $t_0 - t_9$.

Kombinací všech výše uvedených metod a postupů je možné zjistit, jakým způsobem si UNS osvojila určitý typ chování, jak úspěšně si ho osvojila a hlavně jakým způsobem toto chování generuje. Taková zjištění pak experimentátorovi umožňují ověřit si jeho vlastní hypotézy o tom, jak je dané chování realizováno u člověka, nebo ho naopak mohou inspirovat k formulování nových hypotéz, které se potom může pokusit experimentálně ověřit na lidských subjektech.

Jak zde již bylo několikrát uvedeno, konekcionistické výpočetní modely se díky své schopnosti učit se na základě svých vlastních zkušeností často používají jako modely procesu osvojování různých kognitivních funkcí. Jednou z výzkumných oblastí, kde lze této vlastnosti umělých neuronových sítí využít, je vývojová psycholingvistika, které je věnována následující kapitola.

3 Vývojová psycholingvistika

„Jazyk nám prichodí taký prirodzený, že ľahko zabúdame, aký je to zvláštny a zázračný dar. Príslušníci nášho biologického druhu na celom svete modulujú dych do sykavých, bzučivých, škripavých či puavých zvukov a počúvajú iných, ktorí robia to isté. Prirodzene, nekonáme tak iba z lásky ku zvukom, ale preto, lebo jednotlivé zložky týchto zvukov obsahujú informácie o zámeroch osoby, ktorá ich vydáva. Ľudia majú poruke nástroj, ktorý im umožňuje výmenu nepredstaviteľného množstva myšlienok. [...] Aký trik sa skrýva za našou schopnosťou navzájom si zaplňať hlavy toľkými rozdielnymi myšlienkami? ... také triky sú dva, slová a pravidlá.,,

STEVEN PINKER, *Slová a pravidlá*

„Nic není v rozumu, co nebylo dříve ve smyslech.“

JOHN LOCKE, *Esej o lidském rozumu*

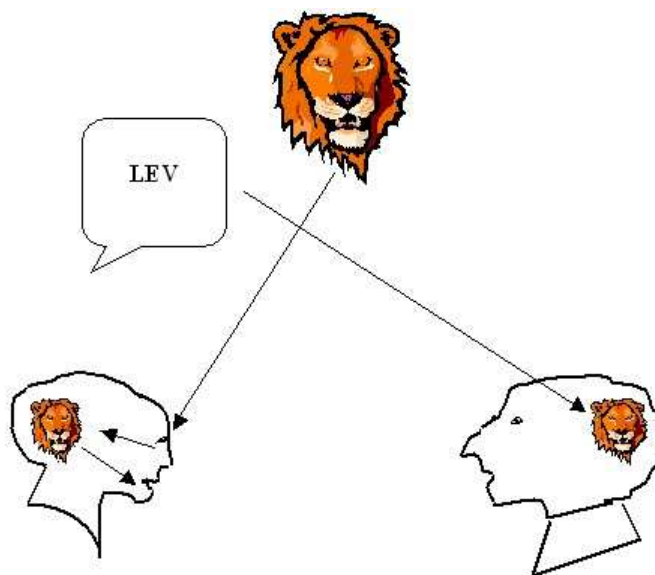
„Nic není v rozumu, co nebylo dříve ve smyslech, kromě rozumu samotného.“

GOTTFRIED WILHELM LEIBNIZ, *Nové úvahy o lidské soudnosti*

Jak napovídá již samotný název, psycholingvistika je vědecká disciplína, která v sobě propojuje poznatky ze dvou různých oborů, a to z psychologie a z lingvistiky. Lingvistika poskytla psycholingvistice základní poznatky o systému jazyka, tj. informace o tom, z jakých prvků se jazyk skládá a jakým způsobem tyto jeho prvky mohou být uspořádány. Psychologie pak dala psycholingvistice do vínku zájem o duševní procesy spojené s produkcí a porozuměním jazyku. Psychologie jazyka je tak tedy o lidské mysli a jejích vnitřních stavech ve stejné míře jako je o samotném jazyce, jak v jedné své práci poznamenal německý psycholog a jeden z prvních průkopníků psycholingvistiky *Wilhelm Wundt* (1900). Charakter moderní psycholingvistiky nezanedbatelným způsobem ovlivnila doba, kdy poprvé spatřila světlo světa. Zrodila se totiž na přelomu 50. a 60. let 20. století, tj. pod vlivem právě se formující („hvězdy“) kognitivní psychologie, která duševní jevy začala pojímat v terminologii mentálních reprezentací a na nich operujících mentálních procesů. Proto je psycholingvistika také často definována jako teoretické a empirické studium mentálních reprezentací a procesů, které stojí v základech schopnosti člověka používat jazyk, a to jak ve smyslu jazykové produkce, tak ve smyslu jeho porozumění. O to, jak se příslušné mentální reprezentace a mentální procesy vyvíjí v průběhu osvojování jazykové kompetence, se pak zajímá vývojová psycholingvistika, která je hlavním předmětem této kapitoly. Dříve než se však dostaneme k některým z výsledků vývojově-psycholingvistického výzkumu, je potřeba zde zmínit alespoň některé z těch nejdůležitějších charakteristik jazyka.

3.1 Základní charakteristiky jazyka

V rámci tradičního a běžně přijímaného *psychologicko-sémiotického* pojetí jazyka (Peregrin, 1999) se o jazyce hovoří jako o **komunikačním systému** či nástroji vyjadřování různých *mentálních obsahů* – myšlenek, pojmů, představ, přesvědčení, přání, záměrů, pocitů apod., resp. jako o nástroji přesouvání těchto mentálních obsahů z jedné mysli do druhé: Výrazy jazyka jsou zde chápány jako *jména* nebo *symbolická označení* (proto *sémiotické*) nějakých *mentálních entit* (proto *psychologicko*). Člověk používající jazyk pak nečiní nic jiného, než že prostřednictvím různých zvuků, gest či klikyháků označuje nějakým arbitrárním, ale celou jazykovou komunitou sdíleným jazykovým výrazem svou myšlenku či jiný obsah své mysli, a umožňuje tak posluchači nebo čtenáři tento mentální obsah pojmut do své vlastní mysli. Jazykovědec *Vladimír Šmilauer* (1966, s. 11-13) takto v úvodu své učebnice české syntaxe píše, že při snaze porozumět funkci jazyka „*vycházíme z předpokladu, že mluvčí i posluchač ovládají touž jazykovou soustavu. Tuto soustavu vytvářejí slova jakožto znaky pro jisté představy (pojmy) a jejich vztahy (jejich prostřednictvím pak pro jisté úseky skutečnosti); slova se spojují do slovních druhů podle toho, označují-li substance, jejich vlastnosti, děje, okolnosti nebo vztahy mezi nimi. [...] Jazykové projevy jsou obsahy vědomí převedené z několikarozměrného obsahu mysli do jedné linie časové a vyjádřené příslušnými*



Obrázek 153: Ilustrace psychologicko-sémiotického pojetí jazyka. Člověk ve své mysli odráží objektivně existujícího lva v podobě nějaké jeho mentální reprezentace či pojmu, který je posléze mluvčím označen prostřednictvím arbitrárního, ale jazykovou komunitou sdíleného jazykového výrazu „lev“. Díky tomu, že na rozdíl od samotného pojmu lva je jeho jazykové označení smyslově dostupné i druhým osobám, umožňuje takový jazykový výraz přenášet mentální obsahy z jedné mysli do druhé.

prostředky jazykové soustavy. [...] U posluchače (čtenáře) je postup opačný: slyší (čte) jazykový projev, ovládá danou jazykovou soustavu [...] vytváří si obsahy vědomí, podobné obsahům vědomí mluvčího.“ Ve stejném duchu – totiž jakožto prostředek vyjadřování idejí, které jsou jinak uzavřeny v lidské mysli – charakterizoval jazyk již *John Locke* ve svém *Eseji o lidském rozumu* (1984), kde píše, že „užitek slov... spočívá v tom, že jsou smyslově vnímatelnými znaky idejí; a ideje, které jsou těmito slovy zastoupeny, jsou jejich vlastním a bezprostředním významem. [...] Slova ve svém prvotním nebo bezprostředním významu nezastupují nic jiného než ideje v mysli toho, kdo těchto slov užívá, bez ohledu na to, jak třeba nedokonale nebo nedbale jsou tyto ideje sejmuty z věcí, které mají představovat.“ Novější „nářez“ dal této představě *Jerry Fodor*, který navrhl dívat se na myšlení jako na druh specifického jazyka (tzv. *jazyk myšlení* nebo také *mentalíza*), který se podobně jako běžný jazyk skládá z určitých základních stavebních jednotek a který má také svoji vlastní „gramatiku“. Rozdíl mezi oběma jazyky spočívá v tom, že zatímco v případě běžného jazyka jsou základními stavebními kameny slova, v případě mentalízy tuto roli hrají různé mentální pojmy či ideje; mentalíza také musí na rozdíl od běžného jazyka disponovat mnohem bohatší sadou „gramatických“ pravidel umožňujících jí se vyvarovat nejrůznějších mnohoznačností tak charakteristických pro běžný jazyk, který ponechává spoustu volného místa pro domýšlení se skutečného významu vyřčené myšlenky.¹ Vyjadřování a přenášení různých mentálních obsahů z jedné mysli do druhé je tak podle Fodora formou překladu z jazyka mentalízy do běžného jazyka (a vice versa), který svoji sémantiku zakládá na původní sémantice vnitřního jazyka myšlení.²

Další důležitou charakteristikou jazyka, která již byla výše naznačena, je že se jedná o **systém arbitrárních znaků**, tzn. že jednotlivé jazykové výrazy, ať už jsou realizovány v jakékoli smyslové modalitě, tedy jako zvuky řeči, gesta rukou nebo klikyháky nějakého písma, nemají žádný přímý vztah k předmětům, ke kterým odkazují. Například slovo „lev“ nevypadá jako lev, nechodí jako lev, ani nevydává zvuky jako lev. Slovo „lev“ označuje lva jenom proto, že jsme se jako malé děti naučili spojovat určitý specifický artikulovaný zvuk

¹ Již v dobách rané historie logiky bylo známo, že věty jsou příliš mnohoznačné na to, aby mohly sloužit jako prostředky přesného usuzování. Z tohoto důvodu byli logici nuceni vytvořit přesnější notační formalismy umožňující jim jemněji rozlišit různé významy, které by věta mohla mít. Výsledkem této snahy jsou různé systémy symbolické logiky jako je například subjekt-predikátový kalkul. Takto například věta „Každý muž miluje nějakou ženu“ může vyjadřovat přinejmenším dva rozdílné významy (jeden, v němž pro každého muže existuje nějaká žena, kterou miluje, a druhý, ve kterém existuje nějaká žena, kterou miluje každý muž), což činí nutným zavést takové syntaktické nástroje jako jsou kvantifikátory nebo závorky, které tyto různé významy umožňují rozlišit.

² Fodorovi se dá vytknout, že dostatečně jednoznačně nespecifikoval, jak získává svou sémantiku samotný jazyk myšlení a že se mu tedy problém sémantiky běžného jazyka nepodařilo vyřešit, ale že ho pouze odsunul o úroveň níže.

s určitým konkrétním významem, resp. s určitým konkrétním mentálním obsahem. Vztah mezi označovaným a označujícím je takto dán čistě na základě konvence či jakési dohody, na kterou člověk „přistupuje“ v průběhu osvojování si daného jazyka. (Určitou výjimkou jsou v tomto ohledu *onomatopoeia* - zvukomalebná slova, jejichž zvuková charakteristika úzce koresponduje s některou z vlastností označovaných objektů.)¹ Když se stejnou „dohodou“ řídí všichni členové daného jazykového společenství, tzn. když všichni používají určitý konkrétní zvuk na vyjádření určitého konkrétního významu, umožňuje to členům takové jazykové komunity si mezi sebou efektivně předávat a vyměňovat myšlenky a pojmy z jedné mysli do druhé. V *sémiotice* (obecné nauce o znacích) se tento vztah často zobrazuje prostřednictvím tzv. *relačního trojúhelníku*, který je založen na *Peircově relační definici* znaku: Jednotlivé vrcholy relačního trojúhelníku tvoří *vehikulum*, *objekt* (označovaný také jako *designát*) a *dispozice*, kde *vehikulum* je smyslově uchopitelná entita, která zastupuje objekt materiální či nemateriální povahy, a to vzhledem ke společné dispozici (schopnosti) mluvčího i adresáta v procesu *semiosis* přiřazovat zastupující *vehikulum* k zastupovanému objektu (Budil, 1998, s. 146-147).

Pro jazyk je dále charakteristické, že v něm lze identifikovat komplexní **hierarchickou strukturu několika různých vrstev** či **rovin**, každou se svojí vlastní třídou základních stavebních prvků - od fonémů, přes morfémy, slova, gramatické kategorie, větné části, věty až po celý diskurz. Na žádné z těchto možných úrovní analýzy přitom není uspořádání jazyka náhodné - na všech se základní prvky řídí **určitým systémem pravidel**, která povolují pouze několik málo možných způsobů uspořádání těchto prvků. S tímto úzce souvisí poslední zde zmíněná výrazná charakteristika jazyka, která spočívá v jeho **generativitě a produktivitě**, tj. v jeho schopnosti z konečné množiny jazykových prvků na základě určitých pravidel generovat prakticky nekonečné množství různých jazykových výpovědí. Nyní se trochu podrobněji podíváme na základní stavební prvky, na které můžeme narazit na jednotlivých hierarchických úrovních analýzy jazyka.

3.1.1 Hláskový systém jazyka

Omezíme-li se na řečovou podobu jazyka, tak na té nejnižší úrovni analýzy se jazyk skládá z **fonémů**, což jsou nejmenší jednotky řečových zvuků, které lze v daném jazyce využít

¹ Již u *Platóna* se lze setkat s úvahami na téma, jakým způsobem jazykové výrazy získávají svůj význam. V jeho dialogu *Kratylos* takto vystupují dvě osoby, *Kratylos* a *Hermogenes*, z nichž každá hájí jednu ze dvou možných pozic: Zatímco *Kratylos* se domnívá, že slova přímočaře odpovídají vlastnostem označovaných předmětů, podle *Hermogena* je význam slov dán čistě na základě dohody mezi lidmi.

k rozlišování významu (Sternberg, 2002). Změna jednoho jediného fonému obvykle mění význam posloupnosti řečových zvuků tvořících nějaké slovo nebo z dané posloupnosti zvuků vytvoří slovo bez významu. Například zvuk /p/ před zvukem /uk/ vytváří slovo /puk/, zatímco zvuk /l/ před zvukem /uk/ vytváří slovo /luk/, tzn. že zvuky /p/ a /l/ jsou dva různé fonémy, které v českém jazyce umožňují rozlišovat významy slov.

Z obrovského množství různých řečových zvuků (**hlásek** neboli **fónů**), které lidská **mluvidla** (ústní dutina, nosní dutina, jazyk, rty, hrtan, měkké patro, čípek atd.) dokáží vygenerovat, každý jazyk využívá jako své fonémy pouze jejich určitou omezenou podskupinu. Hláskový systém českého jazyka je například tvořen pouze 35 různými fonémy - 25 souhláskovými a 10 samohláskovými (5 dlouhými a 5 krátkými). Takže k tomu, aby mluvčí či posluchač českého jazyka mohli plně využít jeho „denotujícího potenciálu“, musí být oba schopni identifikovat rozdíly mezi 35 různými hláskami, které v češtině plní fonémickou úlohu (spočívající v rozlišování významu). Všechny ostatní hlásky jsou z tohoto hlediska irelevantní a jsou mluvčím i posluchačem vnímány pouze jako zvukové varianty některého z 35 fonémů. Hlásky tohoto druhu se také někdy nazývají **alofóny**.

Je tedy zřejmé, že mluvčí každého jazyka se musí ze všech existujících rozdílů mezi všemi možnými hláskami naučit identifikovat právě a jen ty, které v daném jazyce hrají fonémickou úlohu, tzn. ty, které jsou důležité pro rozlišování významu. Přitom platí, že zatímco v jednom jazyce je rozdíl mezi dvěma určitými hláskami významný, v druhém jazyce může být ten samý rozdíl z hlediska rozlišování významu naprosto irelevantní. Známým příkladem jsou hlásky /r/ a /l/, které hrají fonémickou roli v angličtině, nikoli však již v japonštině. Důsledkem je pak to, že zatímco Američané dokáží bez nějakých větších obtíží mezi těmito dvěma různými hláskami rozlišovat, pro Japonce se jedná o jednu a tu samou hlásku, neboť objektivně existující rozdíl mezi hláskami /r/ a /l/ v japonštině nehraje žádnou významotvornou roli. Jinak řečeno, zatímco pro Američany je rozdíl mezi těmito dvěma hláskami fonémický, pro Japonce se jedná o rozdíl alofonní. Každý jazyk si tedy celé spektrum všech možných řečových zvuků „porcuje“ svým vlastním specifickým způsobem do několika různých fonetických kategorií, jejichž hranice se u různých jazyků mohou vyskytovat na různých místech. Je to de facto obdoba situace, kdy zkušený horský vůdce dokáže rozlišit čtyři různé druhy sněhu, zatímco člověk z města vidí pouze obyčejný sníh; přestože oba disponují srovnatelně citlivým senzorickým aparátem, mozek horského vůdce byl nucen se naučit rozlišovat i ty nejjemnější rozdíly ve vlastnostech sněhu, neboť z hlediska jeho obvyklých činností mají i takové drobné rozdíly svůj význam: Je rozdíl jestli je venku prašan nebo firn, když se snažím navoskovat své lyže tak, aby měly ten správný skluz. Pro

městského člověku naproti tomu takové jemné rozlišování žádný velký význam nemá: To, jaký konkrétní typ sněhu přes večer napadl, nehraje příliš velkou roli při ranním rozhodování, co si vzít na sebe; k tomu bohatě postačí schopnost rozlišit, zda je venku sníh, nebo není.

3.1.2 Morfologický systém jazyka

Vedle svého vlastního fonetického inventáře má každý jazyk také soubor pravidel, která specifikují, jakými způsoby se fonémy mohou kombinovat do **morfémů**, což jsou nejmenší jazykové jednotky, které vytvářejí význam slov a které mohou nést samostatný význam (Sternberg, 2002). Rozlišují se dva základní typy morfémů: 1) Prvním typem jsou tzv. **obsahové** nebo také **lexikální** či **volné morfémy**, které plní především sémantickou funkci, tzn. že nesou samostatný význam, samy o sobě dávají nějaký smysl nebo tvoří významové těžiště nějakého slova. Tento typ morfémů velice zhruba odpovídá kořenům slov. 2) Druhým typem jsou tzv. **funkční**, **gramatické** nebo také **vázané morfémy**, které dávají smysl a význam pouze tehdy, když jsou navázány na jiné morfémy. Tento typ morfémů na jedné straně plní funkci sémantickou - to když dotváří a dodává významové detaily k významům obsahových morfémů (tzv. *slovotvorná funkce*) -, na straně druhé plní také funkci gramatickou, neboť umožňuje ohýbání (flexi) obsahových morfémů, tak aby tyto byly v souladu s gramatickým kontextem tvořeným - v případě substantiv - gramatickými kategoriemi rodu, čísla a pádu a - v případě sloves - gramatickými kategoriemi osoby, čísla, času a způsobu (tzv. *tvarotvorná funkce*). Tomuto typu morfémů odpovídají různé afixy (sufixy, prefixy a infixy, které se nacházejí za, před, resp. uvnitř slovního kořene), koncovky, pomocná slova apod.

3.1.3 Lexikální systém jazyka

Soubor všech morfémů nějakého jazyka nebo nějaké konkrétní osoby se v lingvistice označuje pojmem **lexikon** (Sternberg, 2002). Jeho rozsah se u průměrného dospělého (anglicky mluvícího) člověka může pohybovat zhruba v řádu několika desítek tisíc položek. Stejně jako v případě fonémů, také v případě morfémů má každý jazyk soubor pravidel specifikujících, jakým způsobem se jednotlivé druhy morfémů mohou spojovat a kombinovat. Kombinací jednotlivých položek lexikonu vzniká **slovník**, který představuje veškerou slovní zásobu, kterou má daný jazyk nebo daná osoba k dispozici. Rozsah takového slovníku se

může opět pohybovat v řádu desítek nebo i stovek tisíc položek.¹ Slova obsažená ve slovníku se často dělí na **slova obsahová** a **funkční**, kde obsahová slova ve svém významu odkazují k nějakým více či méně konkrétním objektům a událostem, zatímco slova funkční specifikují vztah mezi slovy obsahovými. Například ve větě *Auto narazilo do svodidel* jsou slova *auto*, *narazilo* a *svodidel* slovy obsahovými a slovo *do* je slovem funkčním. Oprávněnost rozlišování mezi těmito dvěma kategoriemi slov podporují četná klinická pozorování, která dokládají, že mentální reprezentace obsahových a funkčních slov je vázána na odlišné neuronální struktury: Zatímco při poškození tzv. *Brockova centra řeči*² (nacházejícího se v zadní části frontálního laloku levé mozkové hemisféry) mají lidé problémy s používáním funkčních slov, ale nikoli s používáním slov obsahových, při poškození tzv. *Wernickeho centra řeči*³ (nacházejícího se v okolí zadní části Sylviovy rýhy na hranicích temenního a spánkového laloku levé mozkové hemisféry) je tomu přesně naopak (tzv. *dvojí disociace*⁴).

3.1.4 Syntaktický systém jazyka

Další úroveň analýzy jazyka je syntax, kterou tvoří pravidla kombinace jednotlivých slov podle jejich příslušnosti k jednotlivým gramatickým kategoriím do syntagmat⁵, frází⁶ a vět, jejichž celkový význam se dá odvodit z významů jednotlivých slov a ze způsobu jejich uspořádání. Syntaktická pravidla, tj. pravidla slovosledu, kterými se řídí pořadí slov ve větě, jsou řešením problému, jak do lineárního sledu řečových zvuků zakódovat komplexní hierarchickou strukturu pojmů či myšlenek, které se člověk snaží vyjádřit a přenést do mysli druhé osoby. Takto je například zásadní rozdíl mezi dvěma situacemi, kdy po verandě honí

¹ Údaj se opět týká anglického mluvčího. Velikost slovníku je značně ovlivněna bohatostí morfologie jazyka. Zatímco v češtině je takto možné vyjádřit velké množství různých významů či významových odstínů tím, že se ke kořenu slova přidají různé přípony a předpony, v jiných jazycích jsou k vyjádření stejných významů a významových odstínů zapotřebí slova s odlišnými slovními kořeny.

² Důsledkem poškození Brockova centra řeči je tzv. *Brockova (neplynulá) afázie*, kdy pacienti „mluví pomalu, tři až sedm slov za minutu. Jejich řeč je namáhavá..., nejsou schopni tvořit řeč dle gramatických pravidel..., neužívají spojky, předložky, pomocná slovesa, přípony, které určují minulý čas.“ (Koukolík, 2005, s. 84)

³ Důsledkem poškození Wernickeho centra řeči je tzv. *Wernickeho (plynulá) afázie*, kdy pacienti „mluví plynule, někdy tak rychle, že drmolí až sto padesát slov za minutu. Dobře vyslovují. Jejich věty jsou přiměřeně dlouhé a melodické. Špatně však pojmenovávají, špatně mluvenou řeč chápou a opakují. Některá nebo všechna jejich slova bývají nesrozumitelná.“ (Koukolík, 2005, s. 84)

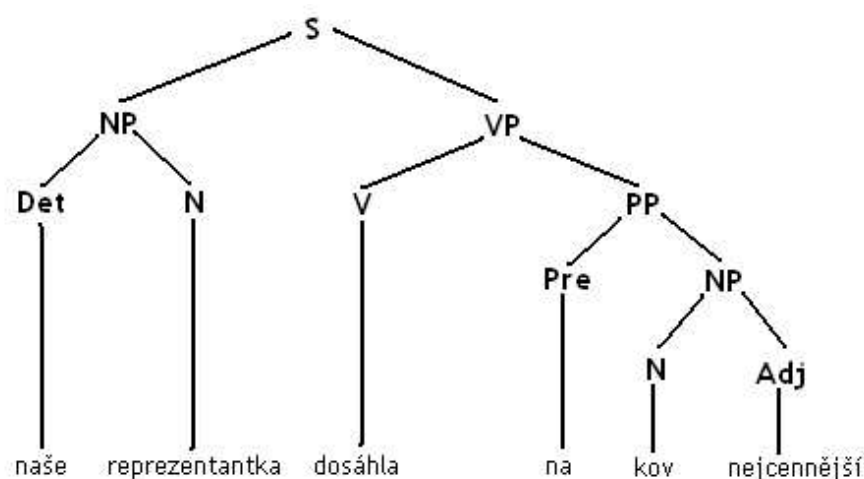
⁴ K dvojité disociaci obecně dochází tehdy, když jedna skupina pacientů dobře zvládá úlohu X, ale má potíže s úlohou Y, a zároveň druhá skupina pacientů vykazuje přesně opačný vzorec chování, tzn. že má potíže se zvládnutím úlohy X, ale dokáže si dobře poradit s úlohou Y. Dvojí disociace je pak považována za důkaz, že řešení úloh X a Y je zprostředkováno dvěma různými a na sobě nezávislými kognitivními procesy či neuronálními strukturami.

⁵ Syntagma – dvojice výrazů těsně spojená vzájemným vztahem (např. *velmi pěkná, rychlé auto* apod.).

⁶ Fráze – seskupení slov, která se ve větě chovají jako jednotka s určitým významem (např. *teší se na Vánoce, chlapec s baseballovou pálkou, na jaře* apod.).

pes kočku a kdy po verandě kočka honí psa. Přestože lze obě situace popsat prostřednictvím stejných slov (kočka, pes, veranda), k tomu, aby bylo možné rozlišit, o jakou konkrétní situaci se jedná – tedy kdo dělá co a komu –, musí mluvčí tato slova určitým specifickým způsobem uspořádat a seřadit (a v případě češtiny – jakožto flektivního jazyka – jim také dát správný tvar). Každý jazyk má přitom svoje vlastní specifická pravidla, jak pojmovou strukturu myšlenek do lineární posloupnosti slov převést a zakódovat. K analýze toho, jakým způsobem jazyk toto kódování provádí, se používají různé metody a nástroje syntaktické analýzy.

Jedním z takových nástrojů syntaktické analýzy je tzv. *frázový ukazatel* (který se používá zejména v rámci americké lingvistické tradice, která je silně ovlivněna *generativní gramatikou Noama Chomskyho*). Frázový ukazatel každou větu dělí do dvou základních částí – na jmennou frázi (*noun phrase*), které v české lingvistické tradici odpovídá podmětová část (obsahující přinejmenším jedno substantivum, případně i nějaký jeho deskriptor či determinátor), a na slovesnou frázi (*verb phrase*), které v české lingvistické tradici odpovídá část přísudková (obsahující minimálně jedno sloveso, případně nějaký jeho objekt). Tyto dvě hlavní větné části se dělí na další dvě části, ty na další dvě atd. dokud se nedospěje k jednotlivým výrazům. Na obrázku 154 je takto zachycena struktura věty *Naše reprezentantka dosáhla na kov nejcennější*. Ze stromového diagramu frázovitého ukazatele je celkem dobře patrné, že za lineárním sledem slov se skrývá poměrně komplexní hierarchická struktura větných členů, které jsou od sebe různě vzdálené a různým způsobem na sobě závislé.



Obrázek 154: Struktura věty „Naše reprezentantka dosáhla na kov nejcennější“ zachycená prostřednictvím frázového ukazatele.

Syntaktická pravidla jsou však pouze jednou ze dvou částí gramatiky jazyka. Ta se vedle (syntaktických) pravidel pro výstavbu celých vět skládá také z (morfologických) pravidel pro výstavbu tvarů jednotlivých slov (viz oddíl „3.1.1.2 Morfologický systém jazyka“). V této souvislosti je zde dobré poznamenat, že moderní (psycho)lingvistika provozovaná v anglosaském prostředí má tendenci se z gramatických jevů soustředit především na syntax, což je do značné míry dáno tím, že angličtina patří mezi jazyky s poměrně chudou morfologií (omezenou de facto pouze na koncovky množného čísla podstatných jmen, 3. osoby jednotného čísla a minulého času sloves) a pevnou větnou skladbou, na kterou je tak navěšena většina gramatických informací nutných k extrakci hierarchické pojmové struktury z lineární posloupnosti slov ve větě. Čeština má na rozdíl od angličtiny poměrně volný slovosled, tzn. že její syntaktická pravidla připouští více různých způsobů uspořádání větných členů. Tato flexibilita větné skladby (syntaxe) českého jazyka je umožněna jeho bohatší morfologickou strukturou, do které lze zakódovat velké množství různých gramatických informací. Na rozdíl od angličtiny se tak gramatika českého jazyka nevyčerpává syntaxí; stejně důležitou roli v ní hraje také gramatická morfologie. Jinými slovy, ve flektivním jazyce s bohatou morfologií, jakým je čeština, při generování smysluplných výpovědí vedle syntaktických kategorií hrají důležitou roli také různé kategorie morfologické související se skloňováním substantiv a adjektiv a s časováním sloves.

Jsou to právě gramatická pravidla, která stojí v základech téměř neomezených výrazových možností jazyka. Podle Pinkera (2002, s. 21-25) jsou z tohoto hlediska důležité především čtyři následující vlastnosti gramatických pravidel:

- 1) Gramatická pravidla jsou *produktivní*: Tím, že pravidla při sestavování různých kombinací slov nespécifikují posloupnosti konkrétních slov, ale pouze posloupnosti slovních druhů, umožňují člověku vytvářet téměř okamžitě nové kombinace slov a ne jen spoléhat na předem namemorovaná slovní spojení.
- 2) Gramatická pravidla mají *symbolickou* či *abstraktní* povahu, tzn. že nepracují s konkrétními slovy, ale se symboly či s proměnnými (ve smyslu různých gramatických kategorií jako je substantivum, sloveso, předmět, přísudek apod.), za které lze dosadit jakékoli konkrétní slovo – asi tak jako v matematické rovnici můžeme za proměnné x, y, z... dosadit jakékoli číslo.
- 3) Gramatika je příkladem *diskrétního kombinatorického systému*, ve kterém se relativně malá množina prvků nechá podle sady pravidel uspořádat do obrovského souboru různých struktur. V případě jazyka to znamená, že člověk používá určitý kód k překladu mezi pořadími slov ve větě a kombinacemi myšlenek, tj. z konečného

množství jednotlivých (diskrétních) prvků (slov) vybírá a kombinuje prvky do rozsáhlejších struktur (frází a vět) s vlastnostmi, které jsou odlišné od vlastností prvků, ze kterých se skládají. Takto je význam věty *Pes honí kočku* odlišný od významu jakéhokoli ze tří slov, která se v této větě nacházejí, a je také odlišný od významu těch samých slov zkombinovaných v opačném pořadí (*Kočka honí psa*). V diskrétním kombinatorickém systému jako je jazyk tak může existovat neomezené množství zcela odlišných kombinací s nekonečnou škálou (významových) vlastností.

- 4) Gramatika je rovněž příkladem *rekurzivního systému*, ve kterém pravidla produkují výrazy, které obsahují i samy sebe. Díky této vlastnosti gramatických pravidel tak člověk může vypustit z úst například následující větu: *Petr se domníval, že poskytl Honzovi informace o tom, že na nástěnce ve škole se objevila zpráva, že profesor Robert Novák prohlásil, že nikdo nemá věřit tvrzení profesorky Hanky Coufalové, že se většina studentů domnívá, že opakovaná ujišťování vlády, že se nebude zavádět placení školného, jsou v rozporu s prohlášením mluvčího vlády, že většina studentů tvrdí, že platit školné by jim nevadilo vzhledem k tomu, že už beztak jsou se studiem spojeny vysokým finanční náklady*. V tomto příkladě věta vždy obsahuje slovesnou frázi, která obsahuje větu, která obsahuje slovesnou frázi, která obsahuje větu, která obsahuje slovesnou frázi, která... atd. ad infinitum. Díky tomuto postupnému vkládání jedné slovních konstrukcí do druhé lze produkovat nekonečné množství vět, které mohou vyjadřovat nekonečné množství různých myšlenek.

3.1.5 Diskurz

Diskurzivní rovina analýzy jazyka se týká pravidel, kterými se řídí řečové výpovědi na vyšší úrovni než jsou jednotlivé věty, tedy například v kontextu jednoho odstavce, článku, celé knihy, konverzace, určité sociální situace apod. (Sternberg, 2002). V této souvislosti se také často hovoří o tzv. **pragmatice** jazyka, která zkoumá, jak člověk používá jazyk v reálných životních kontextech a situacích. Například věta *Můžeš mi to podat?* může v jednom sociálním kontextu znamenat zdvořilou žádost a v jiné sociální situaci zase dotaz na meze schopností druhé osoby atp.

3.1.6 Prozódie

Pojmem **prozódie** se v lingvistice označují ty zvukové charakteristiky mluvené řeči, které se většinou týkají větších jazykových jednotek než jsou hlásky nebo slabiky. Těmito zvukovými charakteristikami jsou především **přízvuk**, **intonace** a **barva hlasu** (Černý, 1998): Přízvuk je založen na odlišení určité slabiky od jiných slabik zvýšením intenzity příslušného zvuku nebo zvýšením jeho tónu (nebo kombinací obojího). Přízvučná slabika má tak v porovnání s nepřízvučnými slabikami vyšší intenzitu a vyšší tón. Nositelem přízvuku může být slovo, tzv. mluvní takt nebo celá věta. V případě větného přízvuku změna intenzity a výšky zvuku mluvčímu umožňuje označit a zvýraznit nejdůležitější část své výpovědi. Nositelem přízvuku může být rovněž tzv. mluvní takt – rytmická jednotka věty, která obsahuje jedno nebo více slov, z nichž pouze jedno nese přízvuk. Věta se takto může skládat z několika mluvních taktů. Na úrovni jednotlivých slov pak přízvuk plní tzv. *deliminativní* nebo *distinktivní funkci*. Deliminativní funkce spočívá ve vyznačování hranic jednotlivých slov. Tuto funkci může přízvuk plnit jen tehdy, když se jedná o *přízvuk stálý*, tzn. že jeho umístění se stabilně vyskytuje na první, poslední nebo předposlední slabice. Příkladem jazyka se stálým přízvukem je čeština (přízvuk na první slabice), francouzština (přízvuk na poslední slabice) nebo polština (přízvuk na předposlední slabice). Distinktivní funkci pak plní *přízvuk pohyblivý*, který umožňuje rozlišovat význam a gramatické kategorie jinak stejných slov. Příkladem jazyka s pohyblivým přízvukem je ruština. Druhou složkou prozodické struktury jazyka je intonace, která spočívá v určitém specifickém vzorci střídání výšek tónů řečových zvuků. Intonace tak vytváří určitou melodii řeči a to jak na úrovni celých vět, tak jednotlivých mluvních taktů, slov a slabik. Například pro tázací věty je charakteristická silně stoupavá intonace směrem ke konci věty. Naopak pro zvolací věty je typická silně klesající intonace. Ryze individuální zvukovou charakteristikou řečového projevu je pak zabarvení hlasu (timbre), které odráží momentální duševní rozpoložení mluvčího, jeho věk, pohlaví, zdravotní stav, fyzickou kondici apod.

3.2 Osvojování jazyka

I ze stručné a značně zjednodušující charakteristiky jazyka, která je uvedena v předchozím oddíle, je zřejmé, že se jedná o mnoháúrovňový systém, který se na všech svých úrovních řídí souborem komplexních pravidel umožňujících jako celek převádět různé mentální obsahy do lineárního sledu jazykových výrazů. Je proto až s podivem, s jakou bravurou a lehkostí si

malé děti dokáží v časovém rozmezí nepřekračujícím standardní dobu vysokoškolského studia úspěšně osvojit pravidla tak komplikované „hry“, kterou je lidský jazyk, a to navíc ve věku, kdy by asi nikoho nenapadlo chtít po dítěti, aby se naučilo hrát například bridž – karetní hru s nesrovnatelně jednoduššími pravidly než jsou ta, která jsou zapotřebí ke zvládnutí jazyka. Dítě se tedy dokáže relativně snadno a rychle naučit svůj mateřský jazyk, a to aniž by muselo tomuto úkolu věnovat nějaké zvláštní úsilí nebo pozornost – jazyk se učí jen tak jakoby mimochodem, „za pochodu“, tedy způsobem naprosto odlišným od toho, jak se děti ve škole učí například zeměpis nebo dějepis. Celé to připomíná spíše způsob, jakým se dítě učí například chodit - učení se (mateřskému) jazyku je tak tedy především procesem intuitivního zvládnutí určité praktické dovednosti, ke které je člověk nějak predisponován, a nikoli procesem učení se nějaké explicitní teorii.

3.2.1 Protosociální chování dítěte, intuitivní rodičovství a „maminkovština“

Vrozené predispozice k osvojování jazyka lze u dítěte pozorovat od prvních okamžiků po jeho narození. Ty se přitom nijak neomezují pouze na čistě řečové aspekty komunikace, ale týkají se také obecnějších **(proto)sociálních** či **(proto)komunikačních dovedností** usnadňujících navazování kontaktu a spouštění sociálních interakcí mezi dítětem a dospělým, které tvoří rámec pro osvojování specificky lidského způsobu řečové komunikace.

Například již několik hodin po svém narození dítě dokáže bezděčně napodobovat jednoduché faciální výrazy dospělého jako je otevření úst, vypláznutí jazyka nebo vyšpulení jazyka (Meltzoff, Moore, 1983), což již předpokládá poměrně komplexní (třebaže nevědomé) mentální reprezentace a procesy, neboť aby novorozenec byl něčeho takového schopen, musí „vědět“, že to, co před sebou vidí, odpovídá jeho vlastní tváři a že pozorované faciální výrazy může na své vlastní tváři vytvořit vysláním „povelů“ směrem ke svalům na své tváři. Vedle této imitační schopnosti, která se týká výrazů tváře druhých osob, novorozenec je také schopen prostřednictvím vlastní faciální exprese komunikovat směrem ke svému okolí své základní emoční stavy jako je strach, nechuť, překvapení apod. Tento svůj vnitřní stav novorozenec dokáže diferencovaně vyjadřovat také prostřednictvím svého pláče, na nějž matka s dobrým vztahem k dítěti dokáže správně (přestože nevědomě a intuitivně) reagovat podle toho, zda novorozenec pláče například kvůli hladu, bolestem apod. Další vrozenou schopností, která usnadňuje navazování kontaktu mezi dítětem a pečujícím dospělým, je tendence novorozenců preferovat mezi vizuálními podněty ty, které připomínají základní konfiguraci lidského obličeje danou vzájemným postavením očí, nosu a úst. Zpočátku se

jedná spíše o preferenci určitých specifických světelných kontrastů, které oči, nos a ústa na tváři vytvářejí. Avšak po zhruba šesti týdnech života jsou již děti schopné rozpoznat lidskou tvář jako takovou a odlišit ji od jiných vizuálních podnětů vyznačujících se podobnou konfigurací světelných kontrastů jako lidský obličej. Zraková percepce novorozence je v tomto ohledu značně limitována nedostatečnou schopností akomodace, takže novorozenec okolní svět vidí značně rozostřeně, bez drobnějších detailů.

Jistá vrozená předpřipravenost k sociální interakci s malým dítětem je dobře patrná rovněž na straně dospělých pečujících osob. Papoušek a Papoušková (1987) v této souvislosti hovoří o tzv. **intuitivním rodičovství**, které zajišťuje, že chování pečující osoby je jemně vyladěno a přizpůsobeno možnostem a schopnostem novorozence, tak aby se toto mohlo stát součástí přiměřené sociální interakce. Toto specifické chování vůči novorozenci je přitom charakteristické nejen pro matky, ale obecně pro všechny dospělé, kteří jsou s dítětem v nějakém druhu interakce; projevy intuitivního rodičovství lze nalézt dokonce již u šestiletých či osmiletých dětí (chlapců i dívek). Mezi typické projevy intuitivního rodičovství patří výraznější faciální exprese, nápadnější pohyby, které jsou součástí gestikulace, déle trvající pohledy do očí dítěte, přibližování obličeje k dítěti na vzdálenost zhruba 30 cm, což je vzdálenost, na kterou dítě vidí v prvních týdnech svého života nejlépe atd.

Z hlediska osvojování jazyka je klíčovým projevem intuitivního rodičovství specifický způsob, jakým pečující osoba na novorozence mluví. V této souvislosti se často hovoří o tzv. „**maminkovštině**“ či **řeči zaměřené na dítě** (*motherese, baby talk, child-directed speech*), která se oproti běžné řeči vyznačuje řadou prozodických i syntaktických zvláštností (Smolík, 1999): Z hlediska prozodie je pro řeč zaměřenou na dítě charakteristické zvýrazňování intonace, tedy zvětšování rozdílů mezi nejnižšími a nejvyššími tóny řečových zvuků, celkově vyšší posazení hlasu, pomalejší rytmus, delší pauzy a odmlky; z hlediska syntaktického se maminkovština vyznačuje větami s jednodušší gramatickou konstrukcí a s menším počtem slov a častým stereotypním opakováním jazykových výrazů (Cooper, Aslin, 1990). Bylo prokázáno, že novorozenci dávají řeči s těmito specifickými prozodickými a syntaktickými charakteristikami přednost před běžnou řečí, kterou mezi sebou komunikují dospělí (Cooper, Aslin, 1990). To naznačuje, že maminkovština plní určitou adaptivní funkci. Ta se podle Fernaldové (1992) v průběhu vývoje dítěte mění: Zatímco na samém počátku vývoje hlas matky pro dítě funguje jako druh nepodmíněného podnětu, který v něm vyvolává libé pocity klidu a bezpečí, později již dítě dokáže z prozodických vzorců v řeči matky vyčíst důležité informace týkající se některých jejích intencí či emocí (jako je kárání, pochvala, zákaz, upoutání a fixace pozornosti dítěte atp.), ale také informace o struktuře jazyka umožňující

například rozčlenit spojitý tok řeči na jednotlivá slova nebo identifikovat syntaktické složky věty (jako je její podmětová a přísudková část).

Z hlediska osvojování jazykové kompetence je také důležité, že matka (či jiná dospělá pečující osoba) má tendenci k novorozenci během sociální interakce a komunikace přistupovat jako k rovnocennému partnerovi, jehož příspěvek je - v jistém slova smyslu - stejně hodnotný jako její vlastní příspěvek, a to navzdory jeho značně omezeným výrazovým schopnostem. Tato rovnocennost se projevuje především tím, že matka směrem k dítěti nejdříve promluví a poté trpělivě čeká na nějakou jeho reakci, potom si matka opět vezme slovo atd. Novorozenec si takto může již ve velice útlém věku osvojit znalost jedné ze základních charakteristik lidské komunikace, která spočívá v pravidelném střídání projevů jejích účastníků.

3.2.2 Fáze osvojování jazyka

Na základě dosavadních zjištění se zdá, lidé ve všech kulturách při osvojování svého mateřského jazyka procházejí zhruba stejnou posloupností fází jazykového vývoje (Langmeier, Krejčířová, 1998; Sternberg, 2002; Altman, 2005):

1) Pomineme-li křik a pláč, kterým novorozenec dokáže pozorné matce „sdělit“ i něco málo o povaze příčin způsobujících jeho pocit diskomfortu, prvním krokem na cestě od nemluvněte k mluvčímu je tzv. **broukání** (*cooing*), které se u dítěte objevuje někdy kolem druhého měsíce věku, tj. na prahu kojeneckého období. Broukání kojenců má povahu jakési vokalizační hry s hlasovým aparátem, během které dítě vytváří pestrou škálu zvuků a při tom zkoumá, jaké hlásky dokáže vyprodukovat. Broukání všech dětí na světě je stejné, tzn. že nijak neodráží specifickou hláskovou strukturu mateřského jazyka, kterému je kojenec ve svém rodinném prostředí vystaven; dokonce i neslyšící děti procházejí fází broukání. To tedy znamená, že kojenci jsou v tomto období schopni tvořit (a také vnímat, viz oddíl „3.2.3.2 Senzitivita k hláskové struktuře jazyka“) i nefonémické hlásky, tj. i ty hlásky, které nejsou součástí jejich (budoucího) mateřského jazyka.

2) Fáze broukání plynule přechází do fáze tzv. **žvatlání** (*babbling*). To se u dětí objevuje někdy okolo pátého šestého měsíce věku. V této fázi jazykového vývoje již mezi produkovánými zvuky převažují (fonémické) hlásky, které jsou specifické pro jazykové prostředí, v němž kojenec vyrůstá. S tímto těsně koresponduje postupně se vytrácející univerzální schopnost kojence rozlišovat hlásky jakéhokoli jazyka – postupně se z něj stává čím dál tím větší specialista na svůj vlastní mateřský jazyk (více viz oddíl „3.2.3.2 Senzitivita

k hláskové struktuře jazyka“). Žvatlání kojenců přitom nekopíruje pouze hláskovou strukturu daného mateřského jazyka, ale odráží také jeho typické intonační vzorce. Díky tomu žvatlající kojenci často budí dojem, že hovoří nějakým svým vlastním jazykem (tzv. *konverzační žvatlání*).

3) Někdy okolo prvního roku věku se u dětí objevují první slova. Existuje zde značná variabilita v tom, kdy dítě takové první slovo vysloví. Zatímco u jednoho dítěte to může být již v 10. měsíci, u jiného dítěte to může být až ve 14. měsíci věku. Ve stejné době děti již většinou rozumějí i několika desítkám, v průměru asi stu různých slov, tzn. že aktivní slovní zásoba dítěte v tomto období značně zaostává za jeho pasivní slovní zásobou, což je skutečnost, kterou lze celkem logicky očekávat vzhledem k tomu, že aktivní používání jazyka již předpokládá určitou minimální úroveň porozumění jazyku a že k jazykové produkci je – na rozdíl jazykové percepce – nezbytná jemná a precizní koordinace jednotlivých komponent řečového aparátu. Období mezi 12. a 18. měsícem života je z hlediska jazykového vývoje **obdobím jednoslovných výpovědí**, tzv. **holofrází**, které děti používají zejména ke sdělování svých potřeb, tužeb, přání, a záměrů. V mentálním slovníku dítěte v tomto období a vlastně až do 3. roku věku zcela převažují substantiva – názvy předmětů a osob, se kterými se dítě často setkává ve svém prostředí (lidé, hračky, jídlo, části těla, dopravní prostředky, zvířata, nástroje atp.). V důsledku nepoměru mezi počtem pojmů, které znají, a počtem slov v jejich mentálním slovníku se děti často dopouštějí tzv. *chyby přílišného rozšiřování (overextension error)* významu slova, kdy jedno slovo používají k označení většího množství příbuzných pojmů. Podle starších výzkumů (Clark, 1973; Nelson, 1973) je tato chyba důsledkem toho, že děti v tomto věku používají k definici objektů příliš málo jejich vymezujících vlastností nebo že tyto definice vytvářejí zejména na základě jejich kauzálního působení či funkce, kterou plní. Takto může dítě například označit čtyři různá zvířata - psa, kočku, křečka a morče - pomocí jednoho slova *pes* proto, že všechna chodí po čtyřech, nebo proto, že všechna plní podobnou funkci domácího mazlíčka. Novější výzkumy (Clark, 1980) však ukazují, že děti si jsou správného významu slov dobře vědomy, neboť dostanou-li za úkol vybrat z obrázků čtyř výše jmenovaných zvířat *psa*, potom neomylně šáhnou po obrázku se psem. Tzn. že děti se chyby přílišného rozšiřování významu slov nemusí dopouštět čistě jenom z důvodu jejich předpojmového myšlení (operujícího s pseudopojmy založenými spíše na vedlejších a nepodstatných charakteristikách objektů), ale také z prostého důvodu snahy dětí překonat prozatímne existující propast mezi počtem pojmů, o kterých by chtěli hovořit, a počtem slov, která k tomu mají k dispozici.

4) V 18. měsíci věku aktivní slovní zásoba dítěte zahrnuje pouze 3-100 slov. Někdy mezi 18. a 24. měsícem věku však dochází k prudkému zvýšení rychlosti nárůstu slovní zásoby, kdy si dítě dokáže osvojit i několik slov denně, takže zatímco v 18 měsících dítě aktivně ovládalo pouze 3 až 100 slov, ve 2 letech to již může být 200 až 300 slov a ve 3 letech již okolo 1000 slov.¹ V této souvislosti se často hovoří přímo o **jazykové** nebo **slovníkové explozi** či **výtrysku** (*naming explosion, vocabulary burst, vocabulary spurt*). Příčina tohoto prudkého nárůstu slovní zásoby se většinou připisuje kvalitativní změně v myšlení dítěte, které právě v tomto období, tj. okolo 2. roku věku, přechází – řečeno v terminologii *Piagetovy* teorie kognitivního vývoje - z fáze senzomotorické inteligence do stadia symbolického myšlení, kdy si dítě začíná uvědomovat, že jedna věc může zastupovat/označovat/reprezentovat věc jinou. Dítě v tomto věku tak začíná skutečně rozumět symbolické povaze slov a uvědomovat si, že každá věc má své jméno: Svět se dětem proměňuje v jedno velké muzeum plné nejrozličnějších exponátů, k nimž je potřeba - za pomoci rodičů a dalšího ochotného „personálu“ - připojit popisky s těmi správnými názvy. Vedle zrychleného nárůstu aktivní slovní zásoby se mezi 18. a 24. měsícem věku objevuje druhý významný vývojový krok – děti v tomto období začínají poprvé **kombinovat slova** a vytvářet dvoj- až trojslovní výpovědi, které se svou jednoduchou gramatikou (bez gramatických či funkčních morfémů, tj. bez předložek, bez členů, bez patřičných pádových koncovek...) podobají spíše strohým telegrafickým sdělením než plnohodnotné lidské konverzaci. Proto se také toto stadium jazykového vývoje označuje jako **období telegrafické řeči**. Poměrně rychle si však děti osvojují základní pravidla gramatiky daného jazyka, takže již někdy okolo čtvrtého pátého roku věku dokáže většina dětí běžně produkovat věty i se složitější gramatickou konstrukcí a 10 let staré děti pak již hovoří jazykem, jehož gramatická struktura je v zásadě totožná s tou, kterou můžeme pozorovat v jazykovém chování dospělých. V následujícím oddíle je podrobněji popsán vývoj některých vybraných kognitivních procesů spojených s osvojováním jazyka.

¹ Tento dramatický nárůst slovní zásoby trvá zhruba až do 15. roku věku, kdy je další přírůstek jen pozvolný a navíc značně závislý na podnětnosti prostředí, ve kterém se člověk pohybuje - zatímco u člověka se základním vzděláním bude mít slovní zásoba tendenci spíše stagnovat, u člověka, který nastoupí dráhu středoškolského nebo vysokoškolského studia, se jeho slovník může i nadále obohacovat o nové výrazy, ovšem již ne v takovém tempu, v jakém docházelo k osvojování nových slov v prvních 15 letech života.

3.2.3 Vývoj některých vybraných kognitivních procesů spojených s osvojováním jazyka¹

3.2.3.1 Senzitivita k prozodické struktuře jazyka

Řada výzkumů prováděných s dětmi mladšími jednoho roku naznačuje, že prvním aspektem řeči, který jsou novorozenci schopni vnímat, je její prozodická struktura.² Mehler a kol. (1988) takto prokázali, že již v prvních dnech po porodu novorozenci preferují svůj mateřský jazyk před cizím jazykem, tzn. že novorozenci již v této době musí být vybaveni nějakým mentálním mechanismem, který jim umožňuje rozlišovat mezi mateřským a jiným, ne-mateřským jazykem. Odfiltrováním vyšších tónů, které posluchači znemožňuje identifikovat jednotlivé hlásky, přitom ale zachovává intonační linii řečového projevu, pak Mehler a kol. prokázali, že mechanismus, který novorozencům umožňuje rozlišovat mezi mateřským a cizím jazykem pracuje s prozodickou a nikoli hláskovou strukturou jazyka.

Přitom se ukazuje, že děti získávají citlivost vůči prozodickým charakteristikám svého mateřského jazyka již v době, kdy se ještě nacházejí v matčině děloze, kde plod může slyšet (zhruba od 25. týdne těhotenství) nejružnější zvuky. Ty mohou mít svůj zdroj jak uvnitř, tak i vně matčina organismu. Asi nejdominantnějším zvukovým podnětem bude vedle tlukotu matčina srdce její hlas. Vzhledem k tomu, že plod je od vnějšího prostředí izolován mnoha vrstvami orgánů a tkání, doléhají k jeho uším pouze zvuky, které jsou zbavené svých vysokofrekvenčních složek, což může vysvětlovat, proč se z mateřského jazyka dítě jako první naučí rozlišovat jeho prozodické a nikoli hláskové charakteristiky – jiné než prozodické aspekty jazyka se zkrátka nemá možnost naučit. Prenatální učení se prozodické struktuře jazyka prokázali ve svém výzkumu DeCasper a Spence (1986), v němž těhotné ženy v průběhu několika posledních týdnů svého těhotenství nahlas předčítaly svým dětem úryvky z nějaké povídky; několik dní po porodu se pak za pomoci techniky nenutritivního sání

¹ Obsah i struktura tohoto oddílu vycházejí především z přehledových studií Filipa Smolíka (2006) a Gerryho Altmanna (2001).

² Vzhledem k tomu, že malé děti disponují pouze omezenými výrazovými, motorickými, volními a dalšími schopnostmi, musí se ke zkoumání jejich kognitivních schopností spojených s řečovou percepcí využívat různé nepřímé metody (například *technika nenutritivního sání*, *poslechové preference* či *podmíněného zrakového vyhledávání*). Ty jsou většinou založeny na **tendenci malých dětí preferovat nové, nebo naopak staré a dobře známé podněty**. V případě, že dítě nějakou takovou preferenci vůči určité třídě podnětů (svým nonverbálním chováním) vykazuje, pak se to bere jako důkaz, že dítě disponuje nějakým mentálním mechanismem, který mu umožňuje mezi danými podněty rozlišovat, a že je tedy vůči některým jejich aspektům senzitivní. Výzkumné techniky založené na preferenci nového využívají **principu habituace**, kdy se dítě systematicky učí nereagovat na opakovaně prezentovaný podnět: Dítěti je nejdříve (v tzv. familiarizační fázi experimentu) opakovaně prezentován určitý podnět a to do chvíle, kdy se o něj dítě přestane zajímat; poté se prezentuje nový podnět; když dítě dál nejeví zájem, lze na základě toho usuzovat, že nový podnět dítě vnímá jako starý podnět, resp. že mezi novým a starým podnětem nečiní žádného rozdílu; naopak v případě, že nový podnět pozornost dítěte upoutá, lze předpokládat, že dítě mezi oběma podněty rozlišovat dokáže (Smolík, 2006).

zjišťovaly zvukové a hlasové preference novorozenců; zjistilo se, že 1) novorozenci dávají přednost předčítání od své matky před předčítáním stejného textu od cizí ženy, tedy že novorozenci preferují hlas své matky před cizím ženským hlasem a 2) že novorozenci dávají přednost těm částem povídky, které měli možnost slyšet ještě v děloze, před těmi úryvky, které takto možnost slyšet neměli

3.2.3.2 Senzitivita k hláskové struktuře jazyka

Novorozenci nevykazují senzitivitu pouze k prozodické struktuře jazyka. Velice brzy jsou také schopni rozlišovat mezi jednotlivými zvuky jazyka – fonémy. Tuto senzitivitu vůči rozdílům mezi hláskami u dětí jako první prokázali Eimase a kol. (1971), v jejichž výzkumu jedno- a čtyřměsíční děti dokázaly rozlišovat mezi hláskami /p/ a /b/, které se liší pouze svou znělostí.

Ten samý výzkum také prokázal, že děti rozlišují hlásky stejným způsobem jako dospělí, tzn. že fonémy vnímají kategoriálně (Liberman a kol., 1957): Přestože akustické charakteristiky jednotlivých hlásek plynule přecházejí jedna v druhou, posluchač toto kontinuum vnímá tak, jako by mezi jednotlivými hláskami byly ostré přechody či hranice, na jejichž dvou stranách se nacházejí dvě různé kategorie hlásek. Rozdíly mezi zvuky v rámci jedné kategorie přitom mohou být větší než mezi dvěma zvuky na opačných stranách kategoriální hranice. Tzn. že lidé na jedné straně ignorují odlišnosti v akustických charakteristikách dvou zvuků, jestliže se tyto nacházejí v rámci stejné hláskové kategorie, a na straně druhé jsou citliví i k těm nejmenším rozdílům mezi dvěma zvuky, jestliže je tento rozdíl umísťuje na opačné strany hranice mezi dvěma hláskovými kategoriemi. Pro posluchače tak v nějakém jazykovém vstupu bude existovat pouze /p/, nebo /b/ a nic mezi tím.

Fungování mentálního mechanismu zodpovědného za kategoriální percepci fonémů si lze celkem dobře představit s pomocí grafu stavového prostoru rekurentní sítě, jejíž úloha spočívá v klasifikaci vstupních podnětů do několika různých kategorií (viz obrázek 144 na straně 249): Specifickým nastavením váhové matice a prahových hodnot jednotlivých uzlů sítě získává její stavový prostor - tvořený souborem atraktorů a jejich basins of attraction – určitý specifický „výškový profil“, který rekurentní síť na jedné straně umožňuje klasifikovat dva podněty, které si jsou ve vstupní vrstvě velice podobné, do dvou různých kategorií, a na straně druhé vnímat dva podněty, které jsou ve vstupní vrstvě kódované jako rozdílné a sobě nepodobné, jako příklady jedné a té samé kategorie.

Eimase a kol. tedy prokázali, že již v prvních měsících svého života děti dokáží v jazykovém vstupu rozpoznávat základní stavební prvky jazyka. Tato skutečnost byla svého času považována za doklad toho, že člověk se již rodí s vrozenou schopností vnímat specifické charakteristiky jazyka, tedy že již při narození je vybaven určitými specializovanými detektory fonémů. Později se však ukázalo, že kategorická percepce se nijak neomezuje pouze na člověka, ale že ji jsou schopni i jiné živočišné druhy (Kuhl, Miller, 1975), a že se ani neomezuje pouze na řečovou doménu, ale že ji lze pozorovat také například při vnímání barev nebo emočních výrazů ve tváři (Jusczyk, 1997).

Důsledkem kategorické percepce fonémů je to, že lidé na jedné straně dokáží velice dobře diferencovat mezi zvuky, které spadají pod různé fonémy jejich jazyka, na straně druhé však jen velice obtížně rozlišují ty zvuky, které v jejich jazyce odpovídají stejnému fonému. Vzhledem k tomu, že každý jazyk vytyčuje hranice fonetických kategorií v kontinuu řečových zvuků trochu odlišným způsobem, je pro lidi často velice obtížné rozlišit ty řečové zvuky cizího jazyka, mezi nimiž je v tomto cizím jazyce rozdíl fonémický zatímco v mateřském jazyce je tento rozdíl pouze alofonní. Zde již jednou uváděným příkladem jsou hlásky /r/ a /l/, které Američan vnímá jako dva odlišné fonémy, zatímco pro rodilého Japonce zvuky odpovídající těmto dvěma hláskám spadají do jedné zvukové kategorie. Na rozdíl od dospělých děti do určitého věku dokáží dobře rozlišovat mezi zvuky odpovídajícími rozdílným fonémům jakéhokoli jazyka, tedy nejen toho mateřského. Tzn. že děti jsou schopné v jazykovém vstupu vnímat i ty fonetické kategorie, vůči kterým jsou jiní než rodilí mluvčí daného jazyka „hluší“. V průběhu prvního roku života se však děti postupně specializují na diskriminaci fonémů svého mateřského jazyka, takže někdy okolo 10. měsíce věku svou původní univerzální diskriminační schopnost ztrácejí a začínají být schopné rozlišovat pouze ty fonetické kategorie, které jsou relevantní z hlediska rozlišování významu v jejich vlastním mateřském jazyce. Jako příklad zde můžeme uvést schopnost sedmiměsíčních dětí z anglicky mluvících rodin rozlišovat mezi fonémy, které se používají pouze v hindštině; jedenáctiměsíční děti z anglofonního prostředí již něčeho takového schopné nejsou (Werker, Tess, 1984).

3.2.3.3 Segmentace spojitého toku řeči

Aby se dítě vůbec mohlo začít učit slova a osvojovat si na nich definovaná gramatická pravidla (v podobě jejich slovosledu a tvarosloví), musí být nejdříve schopné jednotlivá slova vyčlenit z toku zvukových podnětů, které tvoří lidskou řeč, což není nijak jednoduchý úkol,

neboť v mluveném jazyce se na rozdíl od psaného textu nevyskytují žádné mezery nebo čárky, kterých by bylo možné se „chytit“ při stanovování hranic mezi jednotlivými slovy. Řeč fakticky tvoří jeden souvislý proud zvuku a

jenomdíkytréninkukteréhosenám dostalovprvnímrocenašehoživotasenámzdážeposloucháme posloupnostjasněodsebeodlišitelnýchslov

Pro dítě, které tímto tréninkem ještě neprošlo, se však řeč jeví taková, jaká ve skutečnost je, tedy jako jeden souvislý proud zvukových podnětů. Situaci novorozence tak lze celkem dobře připodobnit k situaci kryptoanalytika, který sedí před zašifrovaným textem a snaží se ve zdánlivě zcela náhodné posloupnosti písmen najít nějakou skrytou pravidelnost, kterou by bylo možné využít k překlopení šifrovaného textu v text otevřený. Simon Singh (2003, s. 75) takto luštění šifer přirovnává ke slézání strmé a (zdánlivě) dokonale hladké stěny, ve které se kryptoanalytik pokouší nalézt nějakou škvíru či trhlinku, která by mu mohla posloužit jako opora při jejím slézání, což je přirovnání, které celkem dobře „sedí“ také na situaci malého dítěte pokoušejícího se lokalizovat hranice jednotlivých slov v souvislém toku řeči - také ono se musí pokoušet najít ve zdánlivě zcela náhodné posloupnosti řečových zvuků (hladké stěně) jisté zákonitosti (škvíry a trhliny), které by mu mohly pomoci v identifikaci jednotlivých slov.

Vedle rozpoznání obecného šifrovacího algoritmu je hlavním kryptoanalytickým problémem astronomicky vysoký počet možných klíčů, které lze s pomocí takového šifrovacího algoritmu vygenerovat. Jako příklad zde lze uvést poměrně jednoduchou *monoalfabetickou substituční šifru*, při níž se každé písmeno otevřeného textu napsaného v otevřené abecedě nahradí jiným písmenem šifrové abecedy, čímž vznikne šifrový text, který dokáže přečíst, tj. převést zpět do otevřeného textu, pouze ten, kdo má k dispozici šifrovou abecedu. Známým příkladem monoalfabetické substituční šifry je *Cesarova posunová šifra*, která spočívá v nahrazování každého písmene otevřené zprávy písmenem nacházejícím se v abecedě o tři pozice dále, tzn. že v případě Cesarovy šifry jsou otevřená a šifrová abeceda vůči sobě posunuté o tři pozice. Posunutí otevřené a šifrové abecedy však může být libovolné, tzn. že možných klíčů existuje (při použití české abecedy s 32 znaky) zhruba 260 000 000 000 000 000 000 000 000 (32!). Hledání toho správného klíče je pak doslova hledáním jehly v kupce sena. Jedním ze základních kryptoanalytických nástrojů, který kryptoanalytikovi umožňuje značně zredukovat uvažovanou množinu možných klíčů, je *frekvenční analýza*. Tento nástroj využívá k luštění šifer jistou elementární statistickou zákonitost týkající se četnosti výskytu jednotlivých hlásek (nebo písmen) v daném jazyce – ty

se totiž nevyskytují stejně často (například v českém jazyce se zcela jistě v psaných textech objevuje písmeno *k* mnohem častěji než písmeno *x* nebo *q*): Víme-li, v jakém jazyce je zakódovaná zpráva napsaná, potom stačí na dostatečně reprezentativním vzorku textů spočítat relativní četnosti jednotlivých písmen a to samé potom provést s písmeny obsaženými v šifrovaném textu. Poté se jednoduše spárují písmena z otevřené a šifrové abecedy podle jejich relativního pořadí - první nejčastější písmeno z otevřené abecedy s prvním nejčastějším písmenem ze šifrové abecedy atd. V kombinaci se znalostí kontextu a jen s trochou logického uvažování se pak lze již celkem snadno dobrat otevřeného textu.

Nyní je otázka, zda také malé děti dokáží využívat podobných statistických zákonitostí obsažených v jazykovém vstupu k segmentaci spojitého toku řeči na jednotlivá slova. Současné výzkumy naznačují, že to dokážou. *Saffran, Aslin a Newport* (1996) takto ve svém výzkumu prokázaly, že osmiměsíční děti jsou schopné se během pouze zhruba dvouminutové expozici umělému jazyku sestávajícímu z posloupnosti nesmyslných slabik (např. *bidakupado...*) naučit rozeznávat umělá „slova“ - sestávající z kombinace tří slabik, které se v proudu umělé řeči vyskytovaly vždy pohromadě a ve stejném pořadí - od „ne-slov“ – skládajících se rovněž ze tří slabik, které se ovšem v umělé řeči nenacházely vždy vedle sebe a ve stejném pořadí. Tzn. že děti ve věku 8 měsíců musí disponovat nějakým mentálním mechanismem, který jim umožňuje snadno a rychle analyzovat jednoduché statistické zákonitosti v jazykovém vstupu a používat získané informace jako druh fonotaktického vodítka či nápovědi (*cue*) k segmentaci spojitého toku řeči na jednotlivá slova. Toto zjištění je důležité zejména v souvislosti s probíhající diskusí na téma povahy vrozených předpokladů k osvojení jazyka (viz oddíl „3.2.4.1 Nativismus vs. empirismus“).

Dalším dokladem toho, že děti jsou schopné detekovat elementární statistické zákonitosti výskytu jednotlivých hlásek ve svém mateřském jazyce a využívat je k segmentaci řeči, je výzkum *Mattyse, Jusczyka a kolektivu autorů* (1999), kteří zkoumali, zda devítiměsíční děti (anglicky mluvících rodičů) dokáží diferencovat mezi dvěma různými typy hláskových dvojic – totiž mezi těmi dvojicemi hlásek, které se v anglickém jazyce pohromadě vyskytují téměř výhradně uvnitř jednoho slova, ale téměř nikdy na hranici dvou různých slov, a těmi dvojicemi hlásek, u kterých platí přesný opak. K tomu, aby mohli na tuto otázku odpovědět, využili výsledků dřívějších výzkumů, které ukázaly, že malé děti vyrůstající v anglofonním prostředí preferují „slova“ s přízvukem na první slabice, resp. že k segmentaci spojitého toku řeči využívají (také) prozodického vodítka v podobě umístění přízvuku, který se v angličtině

většinou (v češtině vždy) vyskytuje na první slabice slova.¹ Mattys, Jusczyk a kol. vytvořili umělá slova, která uprostřed obsahovala jeden ze dvou možných typů hláskových dvojic. Poté tato slova dětem prezentovali s různě umístěným přízvukem (na první, nebo na druhé slabice). Zjistili, že děti preferovali ta slova, u kterých fonotaktické a prozodické vodítko nebylo v konfliktu. Takže například v případě přízvuku na druhé slabice děti dávaly přednost slovům s tou hláskovou dvojicí, která se v anglickém jazyce vyskytuje častěji na hranicích dvou slov než uvnitř jednoho slova (a vice versa).

Závěrem můžeme tedy konstatovat, že podle všech dosavadních výzkumů děti k segmentaci spojitého toku řeči na jednotlivá slova využívají jak fonotaktických tak i prozodických vodítek. Tzn. že již někdy od začátku druhé poloviny prvního roku života děti musí ve svém mozku disponovat nějakým druhem statistického procesoru (*statistical learning device*), který jim umožňuje těžit z toho, že ani přízvuk ani jednotlivé hlásky se v řečovém vstupu nevyskytují zcela náhodně, ale že se jejich výskyt řídí určitými pravidly či zákonitostmi, která vytvářejí hláskovou a prozodickou strukturu jejich mateřského jazyka.²

3.2.3.4 Učení se významům slov

Jak již zde bylo uvedeno, hlavní funkce jazyka spočívá v přenášení mentálních obsahů z jedné mysli do druhé. Tento přenos se děje primárně prostřednictvím smyslově zaznamatelných slov, jejichž významem jsou jednotlivé mentální obsahy, ke kterým odkazují. Aby se dítě mohlo stát plnohodnotným členem své jazykové komunity, musí se naučit, jakým způsobem jsou v daném jazyce k sobě přiřazovány jednotlivá slova a jim odpovídající mentální obsahy, resp. předměty a události, jichž jsou mentální obsahy pouhou vnitřní reprezentací. Naučit se významy slov tak představuje jeden ze základních vývojových kroků na cestě od nemluvněte k plně kompetentnímu mluvčímu.

Osvojování významu slova ve své podstatě probíhá na základě docela jednoduché asociace mezi určitým předmětem (či událostí) a zvukem, kterým se k tomuto předmětu v dané jazykové komunitě odkazuje. Jakkoli se může takový proces asociování zdát jednoduchým a přímočarým řešením otázky, jak si dítě osvojuje významy slov, jsou s ním spojeny jisté nezanedbatelné obtíže. Hlavní problém je v tom, že ne vždy je patrné, který zvuk přiřadit ke

¹ V 7,5 měsících (nikoli ale v 6 měsících) děti dokázaly v proudu řeči rozeznat předem prezentovaná jednoslabičná slova (Jusczyk, Aslin, 1995); při prezentaci dvojslabičných slov pak dokázaly rozeznat pouze ta slova, která měla přízvuk na první slabice (Newsome, Jusczyk, 1995).

² Vzhledem k časně senzitivitě novorozenců vůči prozodické struktuře jejich mateřského jazyka (který dokáží bezpečně odlišit od cizích jazyků, které neměli možnost slyšet v matčině děloze), je vysoce pravděpodobné, že nějaký takový statistický procesor je u dítěte v činnosti již nějakou dobu před porodem.

kterému předmětu. V každém okamžiku se totiž dítě nachází v prostředí, které je na jedné straně plné nejrůznějších (řečových) zvuků a na straně druhé zahrnuje také spoustu předmětů a událostí, ke kterým by právě vyřčené slovo mohlo odkazovat. Dítě tak při osvojování významu slov stojí před velice obtížným problémem, jak zjistit, který z těch myriády zvuků odpovídá kterému z těch nespočetných objektů a událostí, které se kolem něj právě nacházejí nebo odehrávají. Takto například nějaká dospělá osoba může říct malému dítěti *Podívej, pejsek honí kočičku*. Přestože se dítě může díky některým svým znalostem (může například již znát význam slova *kočička* a díky určité rudimentární znalosti gramatiky může také vědět, že slovo *pejsek* je substantivum, tedy že odkazuje k nějaké věci a nikoli k ději) dovtípit, že zvuky odpovídající slovu *pejsek* se vztahují k té velké chlupaté a štěkající věci, co se žene za tím, o čem ví, že se tomu říká *kočička*, musí dítě ještě určit, zda se slovo *pejsek* vztahuje na tohoto konkrétního psa, na psy či zvířata obecně, na jeho barvu či tvar, nebo na nějakou jeho dílčí část, například na ocásek nebo na ucho. Možností je tedy bezpočet a dítě mezi nimi musí vybrat tu správnou. (Post)analytický filozof *Willard Van Orman Quine* (1960) v této souvislosti hovoří o tzv. *nevymezitelnosti reference*.¹

Podobně jako kryptoanalytik využívá k redukci obrovského počtu možných klíčů statistických zákonitostí v distribuci jednotlivých hlásek daného jazyka, tak podle některých psycholingvistů také dítě ve své snaze naučit se významům slov vychází z určitých specifických předpokladů, které mu umožňují značně omezit počet objektů a událostí, ke kterým by zaslechnuté slovo mohlo odkazovat. Podle výsledků řady výzkumů děti například předpokládají, že

- substantiva odkazují spíše k celým objektům než k jejich částem nebo k materiálům, ze kterých se skládají (Markman, Hutchinson, 1984);
- substantiva označují objekty mající stejný nebo podobný tvar (Landau, Jones, Smith, 1992; Imai, Gentner, Uchida, 1994)

¹ Prostřednictvím myšlenkového experimentu „radikálního překladatele“ (který překládá nějaký naprosto neznámý a doposud ještě nepřeložený jazyk) Quine (1960) dokládá, že „*to, co radikální překladatel primárně identifikuje, když se v terénu dobírá významů slov... neznámého jazyka, není to, jak jsou slova tohoto jazyka spojena s věcmi, ale jakým způsobem a za jakých okolností jsou užívány věty tohoto jazyka. Ukazuje, že cesta od těchto empirických dat k překladové příručce... není daty jednoznačně určena a je ji tedy možné provést různými způsoby... Tomu Quine říká neurčitost překladu. Podobně, chceme-li výrazy domorodého jazyka chápat jako jména nějakých jednoznačně daných věcí (tak jak to chce... selský rozum), máme opět jistou volnost, jak takovou rekonstrukci provést – jak jednoznačně daný vztah celku jazyka k celku světa roz distribuovat do vztahů mezi nějakými atomy..., které nejsou absolutními stavebními kameny světa, ale částmi, na které svět rozparcelováváme, abychom učinili zadost naší potřebě vidět slova jako spočívající na věcech.*“ (Peregrin, 1998, s. 20-21)

- substantiva označují spíše předměty, které patří do stejné třídy objektů, než předměty, které k sobě patří na základě nějakého vzájemného vztahu (Markman, Hutchinson, 1984)
- každý objekt může být označen pouze jedním substantivem (Markman, Wachtel, 1988).
- Vedle těchto předpokladů hraje důležitou roli při osvojování významů slov také určitá rudimentární znalost gramatiky, resp. syntaxe, která rovněž může napomoci zúžit uvažovaný význam nového slova, neboť v různých syntaktických kontextech se mohou často vyskytovat slova pouze s určitou omezenou skupinou významů či významových odstínů; syntaktický kontext takto může dítěti například napovědět, že nové slovo bude spíše substantivum než sloveso, že půjde o ne/transitivní sloveso nebo o počítatelné, resp. hromadné podstatné jméno atp. (Gleitman, 1990).

Podle Smithe (1999) jsou výše uvedené předpoklady spíše než vůdčími principy lexikálního vývoje pouze vedlejším důsledkem obecných principů asociačního učení, kdy se na základě opakovaného a pravidelného spojování jednoho podnětu s druhým vytváří asociační spoj, takže dítě má pak tendenci v přítomnosti prvního podnětu očekávat také druhý podnět. Například tendence předpokládat, že substantiva označují objekty se stejným nebo podobným tvarem, může vzniknout díky opakovanému a pravidelnému asociování substantivní syntaktické formy se statistickými zákonitostmi („rodinnými podobnostmi“) definujícími členství objektu v kategorii pojmenovávané pomocí daného substantiva. A vzhledem k tomu, že většina prvních slov, kterým se děti učí, odkazuje ke kategoriím objektů majících podobný tvar, má jednoduché asociačního učení automaticky a zákonitě za následek apriorní předpoklad, že substantiva odkazují k objektům se stejným nebo podobným tvarem. Kromě principů asociačního učení svou roli mohou hrát také obecné principy zrakového vnímání. Například celek je většinou výraznější a nápadnější než části, ze kterých se skládá, takže celek pak zákonitě přitahuje větší pozornost než jeho části. Tato obecná zákonitost může vysvětlovat, proč mají děti sklon předpokládat, že substantiva označují spíše celé objekty než jejich části.

Poněkud odlišný přístup k řešení otázky, jak si děti osvojují významy slov, nabízí teorie, které mají svůj základ v principech konekcionistického modelování kognitivních procesů. Tyto teorie chápou znalost významu slova jako znalost jazykových a mimojazykových kontextů, ve kterých se dané slovo vyskytuje a používá, což odpovídá pojetí významu, které můžeme nalézt také například u pozdního Wittgensteina (1993), pro kterého význam slova není nějakou konkrétní entitou, ale spíše určitým ohodnocením použitelnosti tohoto slova

v rámci nějaké *jazykové hry* (tento pohled na podstatu významu je charakteristický pro většinu představitelů tzv. *postanalytické filozofie*; viz také poznámka pod čarou na straně 303; více viz Peregrin, 1998). Ve své práci z roku 1990 *Jeffrey Elman* ukázal, že i poměrně jednoduchá umělá neuronová síť dokáže na základě statistické analýzy výskytu jednotlivých slov v jazykovém kontextu (tvořeném dalšími slovy ve větách zjednodušené verze anglického jazyka) identifikovat různé sémantické kategorie (typu živé bytosti vs. neživé objekty, jedlé vs. nejedlé předměty, lidské bytosti vs. živočichové apod.). To, že je něco takového možné, je dáno tím, že jednotlivá slova se v toku řeči nevyskytují zcela náhodně, ale že slova s podobným významem mají tendenci se vyskytovat v podobných (jazykových) kontextech, tedy v sousedství podobných slov. Neuronová síť pak tuto nenáhodnost v řečovém vstupu dokáže objevit, zakódovat do vážených spojů, resp. do míry aktivace uzlů ve skryté vrstvě, a využít ji ke splnění své úlohy (která spočívá v predikci každého dalšího následujícího slova v řečovém vstupu). Zatímco Elman svůj model vytvořil především za účelem simulace osvojování některých gramatických kategorií (které lze v řečovém vstupu objevit na základě stejného principu jako sémantické kategorie), Burgess a Lund (1997) svůj výpočetní model vytvořili s cílem modelovat pouze osvojování významu slov. Podobně jako Elmanovi se jim podařilo prokázat, že na základě statistické analýzy spoluvýskytu jednotlivých slov ve větách lze v jazyce objevit různé dimenze podobnosti, podél kterých jsou rozložena slova spadající do stejné sémantické kategorie.

3.2.3.5 Osvojování gramatických pravidel

Dalším významným krokem v jazykovém vývoji dítěte je osvojení gramatiky – souboru konvenčních pravidel, která specifikují, jakým způsobem je v daném jazyce možné vytvářet smysluplné výpovědi prostřednictvím kombinace (syntaxe či slovosledu) a flexe (skloňování a časování) jednotlivých slov. Například v anglickém jazyce platí syntaktické pravidlo, že větný člen nacházející se před slovesem odpovídá osobě či předmětu, který je aktivním aktérem činnosti popisované slovesem, a že větný člen stojící za slovesem odpovídá osobě nebo předmětu, který je naopak trpným objektem slovesem popisované činnosti. Jinými slovy, v anglickém jazyce platí pravidlo, že gramatická kategorie subjektu (podmětu) se nachází před slovesem a gramatická kategorie objektu (předmětu) za slovesem. Jednoduchou změnou slovosledu tak lze v angličtině poměrně zásadně změnit význam nějakého sdělení. Takto by například český (a anglicky hovořící) fanoušek ledního hokeje zcela jistě potvrdil, že je velký rozdíl mezi větou *The Czech ice hockey players defeated the ice hockey team of*

Canada a větou *The ice hockey team of Canada defeated the Czech ice hockey players* - zatímco v prvním případě má věta za následek vlnu euforie a bujaré oslavy, v druhém případě věta spouští kaskádu negativních emocí a společensky nežádoucího chování frustrovaných jedinců. V češtině jakožto flektivním jazyce s bohatou morfologií jsou gramatické kategorie (předmět, podmět, přísudek...) vyjadřovány nejen prostřednictvím jejich slovosledu (který je v češtině poměrně volný), ale také prostřednictvím tvarosloví. Díky tomu lze v češtině ve většině případů snadno poznat, kdo udělal co a komu (tj. kdo je subjektem a kdo objektem popisované činnosti) i navzdory jejímu značně volnému slovosledu. Například ve větách *Pes honí kočku* a *Kočku honí pes*, které se obě skládají ze stejných slov, ale v zrcadlově převráceném slovosledu, lze s pomocí pádových koncovek bezpečně určit, že to je pes, kdo honí kočku, a nikoli kočka, která honí psa; naproti tomu v angličtině by takové převrácení slovosledu mělo za následek také prohození rolí pronásledovatele a pronásledovaného (*The dog chases the cat* vs. *The cat chases the dog*).

Dítě, které se snaží naučit gramatice svého mateřského jazyka, musí být schopné vyřešit dva základní typy problémů: 1) První problém spočívá v tom, že dítě si musí osvojit znalost základních jazykových kategorií typu slovních druhů (substantiva, slovesa...) nebo větných členů (podmět, předmět, predikát...). 2) Druhým velkým problémem je pak osvojení si gramatických pravidel, která jsou definována na výše zmíněných gramatických kategoriích (viz oddíl „3.1.4 Syntaktický systém jazyka,“). Dítě tak musí vyřešit problém, který je variací na téma *Co bylo dřív, slepice nebo vejce?* Stejně jako v případě slepice a vejce existence jednoho předpokládá existenci toho druhého, tak i v případě gramatických pravidel a kategorií znalost jednoho předpokládá znalost toho druhého: Když dítě nezná gramatické kategorie jednotlivých slov, nemůže se pokusit z dostupných příkladů odvodit gramatická pravidla, která jsou na nich definována; a když dítě nezná gramatická pravidla, potom nemůže využít existujících syntakticko-morfologických zákonitostí v řečovém vstupu k vyvození gramatických kategorií jednotlivých slov (tzv. *problém naučitelnosti, learnability problem*). Podle Altmana (2001) v současné době existují v zásadě tři možné přístupy k řešení tohoto problému, z nichž jeden si bere na pomoc vrozené znalosti některých gramatických jevů, druhý interakci mezi zpracováváním jazykového vstupu a vnímáním kauzální struktury popisovaných událostí a třetí pak statistické zákonitosti ve výskytu jednotlivých slov v jazykovém kontextu.

3.2.3.5.1 Sémantický bootstrapping

Příkladem prvního řešení je Pinkerův **sémantický bootstrapping** (Pinker, 1984, 1987). Podle Pinkera se člověk může gramatice úspěšně naučit jen proto, že již při narození disponuje určitými vrozenými znalostmi o některých univerzálních vlastnostech jazyka (tj. těch vlastnostech, které jsou společné všem lidským jazykům). V tomto je Pinker věrným žákem Noama Chomskyho (1980), který předpokládá, že k tomu, aby se člověk dokázal naučit gramatice svého mateřského jazyka, musí disponovat určitým specializovaným *mentálním orgánem* – tzv. *language acquisition device* -, který si dokáže gramatická pravidla osvojit navzdory ne zrovna ideálním učebním podmínkám (tzv. *argument chudé stimulace*, *poverty of the stimulus argument*). Chomsky (1975) v této souvislosti také hovoří o tzv. *univerzální gramatice* (*Universal Grammar*), která představuje soubor vrozených jazykových principů, jejichž existenci je potřeba předpokládat, abychom mohli vysvětlit ty vlastnosti jazyka, o kterých se nelze domnívat, že by se je člověk mohl sám naučit (viz také oddíl „3.2.4.1 Nativismus vs. empirismus“). Podle Pinkera tyto vrozené znalosti mají jen velice hrubou a povšechnou povahu a týkají se především sémantických kategorií slov, se kterými se člověk může v jazyce setkat (například substantiva jakožto slova označující původce nějakého děje nebo naopak jeho objekt), a gramatických rolí, která tato slova mohou v jazyce plnit (podmět, predikát, předmět...). Součástí vrozené výbavy dítěte je také znalost toho, jak spolu souvisí sémantické kategorie na straně jedné a gramatické kategorie na straně druhé (například že slovo odpovídající substantivu může hrát roli podmětu nebo předmětu). Osvojování gramatiky daného konkrétního jazyka pak probíhá tak, že dítě nejdříve na základě významu nějakého slova (nevědomě) určí jeho sémantickou kategorii a s její pomocí pak (stejně nevědomě) určí jeho odpovídající gramatickou kategorii. Dítě například díky svým vrozeným znalostem jazyka ví, že aktivní činitel nějakého děje odpovídá podmětu věty; jestliže pak dítě bude pozorovat baseballistu odpalujícího míček a jeho rodič to bude komentovat slovy *Baseballista odpaluje míček* a dítě přitom bude vědět, že slovo *baseballista* označuje hráče baseballu, potom si z těchto informací dokáže celkem snadno odvodit, že slovo *baseballista* odpovídá podmětu věty. Navíc získá důležitou informaci, že podmět se v jeho mateřském jazyce může nacházet na začátku věty. Když dítě tímto způsobem nashromáždí dostatečné množství informací o gramatických rolích různých slov, může je použít ke statistické analýze distribuce jednotlivých gramatických kategorií. Dítě vyrůstající v anglofonním prostředí se tak může naučit gramatickému pravidlu, které říká, že v angličtině se podmět většinou nachází na začátku věty atp. Znalost těchto pravidel pak slouží k vyvození

gramatických rolí dalších slov, na jejichž určení velice hrubé vrozené znalosti nestačí. Vrozené znalosti některých gramatických jevů tak v Pinkerově sémantickém bootstrappingu hrají roli jakýchsi krystalických jader, na která se nabalují další gramatické znalosti - jsou tím kritickým prvkem, který systém vychyluje z rovnovážného bodu a umožňuje nastartovat proces osvojování gramatiky mateřského jazyka. Na konci celého procesu je pak kompetentní mluvčí, který je bez nějakého velkého a vědomého úsilí schopen provádět analýzu syntaktických struktur zakódovaných do lineárního sledu slov (tzv. *parsing*), tzn. že snadno a rychle dokáže určit syntaktické kategorie jednotlivých slov, pospojovat je do hierarchické struktury podobné té zachycené ve frázovém diagramu na obrázku 154 na straně 288 a na základě konkrétních informací obsažených ve větě pak určit, kdo udělal co a komu.

3.2.3.5.2 Syntaktický bootstrapping

S poněkud odlišným návrhem řešení problému naučitelnosti gramatiky mateřského jazyka přišla Lila Gleitmanová (1990). Podle ní proces osvojování gramatiky probíhá tak, že dítě na jedné straně využívá percipované kauzální struktury událostí (jako kdo dělá co a komu) k interpretaci syntaktické struktury vět, které tyto události popisují, a na straně druhé tyto větné struktury využívá jako vodítka k řízení své pozornosti při percipování kauzální struktury událostí. Takto například může nějaké dítě upřeně pozorovat kočku jak pije mléko; jeho rodič to může zpozorovat a celou situaci okomentovat slovy *Kočka pije mléko*. Dítě, které bude znát význam slov *kočka* a *mléko* pak dokáže celkem snadno odvodit, že rodičem vyřčená věta se vztahuje k události či ději, kterého se účastní jak kočka tak mléko. A protože má dítě pozorovanou scénu přímo před sebou, může si snadno vyvodit význam slovesa *pít*. Vedle toho si však může vyvodit také jedno důležité gramatické pravidlo, totiž že subjekt (původce děje) se ve větě (alespoň tedy v angličtině) nachází před slovesem a objekt (vůči kterému je děj směřován) naopak za slovesem. Tento mechanismus tak dokáže obstatat dvě důležité věci najednou: Jednak dítěti umožňuje se naučit významy sloves, které bezprostředně nekorespondují s nějakou entitou ve vnějším prostředí¹, a jednak mu umožňuje se něco dozvědět o gramatické struktuře svého mateřského jazyka.

Gleitmanové návrh se ve skutečnosti příliš neliší od toho, se kterým přišel Steven Pinker, neboť jak v případě sémantického tak i v případě syntaktického bootstrappingu dochází

¹ K nastartování tohoto procesu je zapotřebí, aby dítě již mělo k dispozici určitou minimální slovní zásobu sestávající ze substantiv označujících možné aktéry událostí a dějů popisovaných slovesy. Tento předpoklad je také v souladu s empirickým pozorováním, že první slova, kterým se děti učí, jsou substantiva označující běžné předměty z jejich bezprostředního okolí (viz oddíl „3.2.2 Fáze osvojování jazyka,,“).

k interakci sémantických a syntaktických informací. Hlavní rozdíl mezi oběma návrhy spočívá v povaze vrozených dispozic, které Pinker a Gleitmanová předpokládají na straně dítěte, aby se toto mohlo naučit gramatickým pravidlům. Zatímco podle Pinkera se tyto vrozené dispozice týkají specificky jazykových znalostí některých gramatických jevů, podle Gleitmanové je dítěti vrozena spíše obecnější (tj. jazykově nespecifická) schopnost nacházet korelace mezi dvěma různými strukturami - cosi na způsob analogického myšlení, kdy je člověk schopen vnímat dvě různé situace jako stejné či sobě podobné na určité abstraktnější rovině vztahů mezi jejich konstitutivními prvky, které se přitom mohou samy o sobě - napříč dvěma různými situacemi - zásadním způsobem lišit¹

3.2.3.5.3 Distribuční analýza

Další možný způsob, jak vyřešit problém naučitelnosti gramatických pravidel, spočívá ve využívání statistických zákonitostí v distribuci slov v jazykovém kontextu (Karmiloff-Smith, 1979). Podle této teorie dítě (nevědomě) provádí statistickou analýzu distribuce jednotlivých slov v jazykovém kontextu tvořeném všemi zaslechnutými větami, resp. slovy, ze kterých se tyto věty skládají. Následně dítě provádí cosi na způsob shlukové analýzy nebo multidimenzionálního škálování, kdy v jazykovém vstupu hledá skupiny (shluky) slov s podobnými distribučními vlastnostmi, tj. slova, která mají tendenci se vyskytovat ve stejných jazykových kontextech. A vzhledem k tomu, že slova se ve větách daného jazyka nevyskytují náhodně, ale podle určitých gramatických zákonitostí a pravidel, která jsou definována na gramatických kategoriích, je potom vysoce pravděpodobné, že dítětem objevené shluky podobných slov budou odpovídat známým kategoriím jako jsou substantiva, slovesa, podmět, předmět apod. Podle Pinkera (1995) není možné, aby osvojování gramatiky bylo založeno čistě jenom na distribuční analýze výskytu jednotlivých slov v jazykovém kontextu, neboť v jazykovém vstupu podle něj existuje až příliš mnoho statistických a distribučních zákonitostí, z nichž žádné nemají vztah ke gramaticky relevantním kategoriím. Tzn. že podle Pinkera by dítě muselo být schopné nějak určit, které z těch myriády možných statistických zákonitostí vyskytujících se v jeho řečovém vstupu jsou užitečné k odhalení gramatických pravidel jeho mateřského jazyka.

¹ Přestože měla válka Američanů ve Vietnamu a Rusů v Afghánistánu odlišné aktéry a přestože se tyto válečné konflikty odehrávaly v různých lokalitách a také v různé době, na abstraktnější úrovni popisu tyto dva konflikty sdílejí určitou společnou strukturu vztahů mezi jejich vlastními konstitutivními prvky, což tyto dva konflikty činí (v určitém smyslu) totožnými – v obou dvou případech dokázala malá, špatně vyzbrojená, ale o to více odhodlaná, dobře organizovaná a také druhou supervelmocí podporovaná armáda povstalců způsobit výrazné ztráty armádě supervelmoci mající jinak výraznou technologickou a materiální převahu.

3.2.3.5.4 Konekcionalistický model distribuční analýzy

To, že by to pro dítě nemusel být až zas tak velký problém, naznačuje Elmanův úspěšný pokus naučit umělou neuronovou síť rozlišovat některé sémantické a gramatické kategorie čistě na základě prezentace spojitého toku slov tvořících dvoj- nebo trojslovné věty zjednodušené verze anglického jazyka (Elman, 1990). Elman k tomu použil jednoduchou rekurentní síť, která dokáže implementovat funkce, které nezávisí pouze na aktuálním vstupu, ale také na dřívější historii vstupů (více viz oddíl „2.2.2.3.2 Jednoduché rekurentní sítě,,). Tato umělá neuronová síť byla vystavena souboru deseti tisíc jednoduchých, dvoj- až trojslovných vět a to tím způsobem, že v každý časový okamžik t bylo neuronové síti na její vstupní vrstvě prezentováno právě jedno slovo. Věty se skládaly se souboru 29 různých slov - substantiv a sloves -, které bylo možné rozdělit do několika různých (většinou) sémantických kategorií (viz tabulka 10). Tuto informaci však neuronová síť neměla k dispozici v explicitní podobě, neboť jednotlivá slova byla na vstupní vrstvě kódována prostřednictvím lokální (nebo

Category	Examples
NOUN-HUM	man, woman
NOUN-ANIM	cat, mouse
NOUN-INANIM	book, rock
NOUN-AGRESS	dragon, monster
NOUN-FRAG	glass, plate
NOUN-FOOD	cookie, break
VERB-INTRAN	think, sleep
VERB-TRAN	see, chase
VERB-AGPAT	move, break
VERB-PERCEPT	smell, see
VERB-DESTROY	break, smash
VERB-EAT	eat

Tabulka 10: Kategorie lexikálních jednotek použitých v Elmanově simulaci. (převzato z Elman, 1990)

také ortogonální) reprezentace, kdy každému slovu bylo zcela náhodně přiděleno jedno číslo ve vstupním aktivačním vektoru, resp. jeden uzel ve vstupní (a výstupní) vrstvě. Na obrázku 155 je zobrazena posloupnost několika aktivačních vzorců reprezentujících různá slova tak jak jdou za sebou v rámci několika krátkých vět. Z popisu (a zobrazení) reprezentačního schématu by mělo být zřejmé, že neuronová síť neměla k dispozici žádné explicitní informace týkající se gramatických, morfologických nebo prozodických charakteristik jednotlivých slov. Úkolem neuronové sítě bylo se naučit anticipovat svůj příští vstup, resp. produkovat ve své

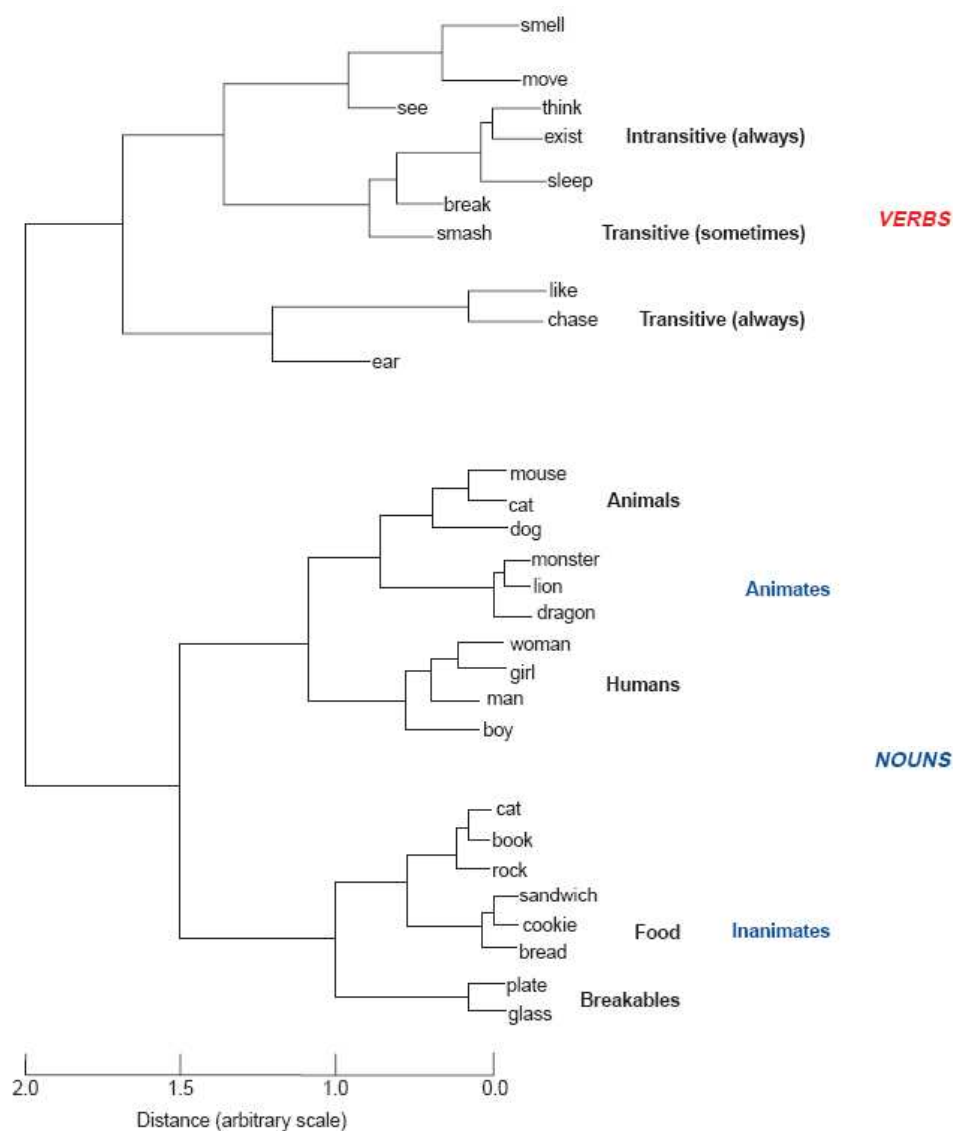
Jak již zde bylo uvedeno, UNS neměla dopředu k dispozici žádné explicitní informace o gramatické nebo sémantické roli jednotlivých slov. Přesto se nakonec naučila anticipovat gramaticky a sémanticky vyhovující skupiny slov jako kdyby tyto informace k dispozici měla. Je tedy zřejmé, že UNS si musela tyto informace nějak osvojit v průběhu svého tréninku. A protože k dispozici měla pouze informaci o distribuci jednotlivých slov, znamená to, že tyto informace mohla vyvodit jedině na základě informací o tom, v kontextu jakých dalších slov se jednotlivá slova obvykle vyskytují, což není až tak překvapivé, víme-li, že slovosled není vůbec náhodný, ale že se řídí určitými syntaktickými zákonitostmi, které jsou definovány na různých gramatických kategoriích. Tzn. že ve slovosledu jsou implicitně zakódovány informace o tom, do jaké gramatické kategorie jednotlivá slova patří. A UNS tyto informace dokáže z posloupnosti slov vyextrahovat: „*Vysokoúrovňové teoretické pojmy jako je gramatická třída slova (například substantivum nebo sloveso) vystávají v rámci systému, který je citlivý pouze ke statistické distribuci slov v rámci vět... Tato emergence [vysokoúrovňových gramatických pojmů] není o moc víc než jednoduché statistické shlukování; vnitřní reprezentace slov, které mají tendenci se objevovat v podobných distribučních kontextech, se shlukují dohromady, a protože substantiva mají tendenci se objevovat v určitých specifických větných kontextech a slovesa v jiných, shlukování slov do těchto dvou tříd (a dalších s ještě jemnějšími rozdíly mezi třídami slov) je tak v určitém slova smyslu statisticky nevyhnutelné.*“¹ (Altmann, 2002)

To, že neuronová síť své chování opravdu zakládá na znalosti gramatických a sémantických rolí jednotlivých slov, lze doložit prostřednictvím shlukové analýzy jejího vnitřního reprezentačního prostoru definovaného vzorci aktivace ve skryté vrstvě vyvolanými prezentací jednotlivých slov na vstupní vrstvě. Způsob, jakým jsou jednotlivá slova v tomto vnitřním reprezentačním prostoru rozmístěna (tj. jak blízko sebe se nachází), vypovídá o tom, jak UNS těmto slovům rozumí – ve smyslu jejich ne/podobnosti z hlediska určitého kritéria, které se UNS jeví být užitečné pro řešení její anticipační úlohy (viz oddíl „2.3.3 Analytická fáze,“). Důležité je, že UNS takto do svého reprezentačního prostoru kóduje pouze ty zákonitosti v jejím jazykovém vstupu, které jsou užitečné pro splnění jejího úkolu spočívajícího v predikci následujícího slova. To má za následek, že ať už se v jazykovém vstupu nacházejí jakékoli statistické zákonitosti, UNS bude ignorovat všechny ty, které nijak

¹ „...higher-level theoretical constructs such as the grammatical class of a word (as noun or verb, for example) would emerge within a system that was sensitive only to the statistical distributions of words within sentences. Importantly, this emergence represented little more than simple statistical clustering; the internal representations of words that would tend to occur in similar distributional contexts would cluster together, and because nouns tend to occur in particular sentential contexts and verbs in others, the clustering of words into these two classes (and others with even finer distinctions between the classes) was in some sense a statistical inevitability.“

nepřispívají ke zdárnému splnění její úlohy (tj. k minimalizaci chyby na její výstupní vrstvě), čímž se UNS zcela automaticky a jakoby jen tak mimochodem vyhne tomu, čeho se obával Pinker (1995), totiž že „statistický procesor“ dítěte bude zahlcen velkým množstvím gramaticky zcela irelevantních statistických zákonitostí v jeho jazykové vstupu. Takže vzhledem k tomu, že to, jaké slovo bude následovat, je determinováno hlavně tím, do jaké sémantické a gramatické kategorie spadají předcházející slova, měla by si UNS svůj vnitřní reprezentační prostor rozvrhnout a rozdělit způsobem odrážejícím především tyto konkrétní charakteristiky jednotlivých slov – tedy alespoň v případě, že v jejím jazykovém vstupu existuje nějaké vodítko (ve smyslu určité zákonitosti), kterého by se mohla při tvarování svého „mentálního prostoru“ chytnout. Pro Elmanovu síť byla tímto vodítkem nenáhodná distribuce jednotlivých slov v jazykovém kontextu dalších slov. A Elmanova UNS si skutečně také dokázala na základě této informace osvojit některé sémantické a gramatické kategorie implicitně zakódované do slovosledu jednotlivých slov. Na obrázku 156 je zachycen dendrogram hierarchické shlukové analýzy vnitřního reprezentačního prostoru Elmanovy UNS. Z dendrogramu je dobře patrné, že vzorce aktivace vyvolané ve skryté vrstvě prezentací jednotlivých slov odráží příslušnost těchto slov do několika různých gramatických a sémantických kategorií jako jsou slovesa/substantiva, ne/tranzitivní slovesa, živé/neživé objekty, živočichové/lidské bytosti atp. Tzn. že UNS se (díky specifickému nastavení vah spojujících vyladěných v průběhu jejího tréninku) podařilo ve své skryté vrstvě přetransformovat původně zcela náhodné kódování jednotlivých slov (ve vstupní vrstvě) takovým způsobem, aby jejich reprezentace (ve skryté vrstvě) systematicky odrážela jejich gramatickou a sémantickou roli, kterou hráli ve větách, kterým byla UNS během tréninku vystavena.

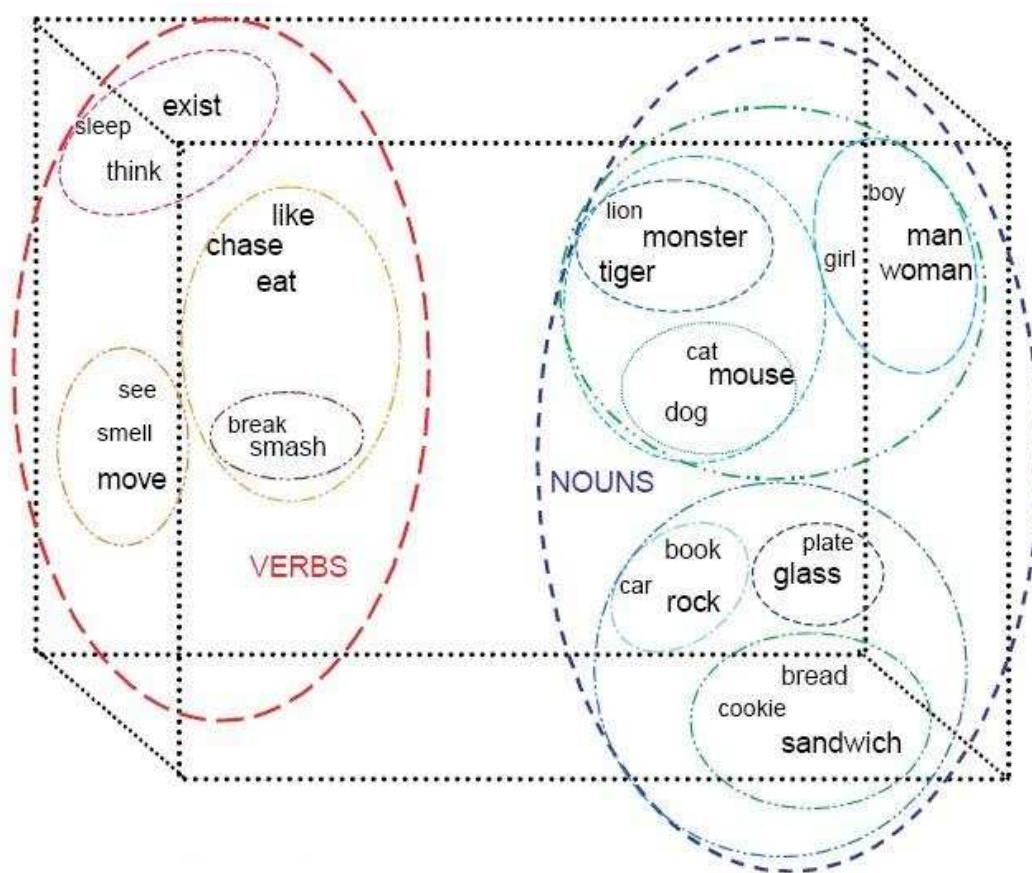
Příslušnost slova do určité kategorie je dána oblastí multidimenzionálního reprezentačního prostoru, ve kterém se dané slovo nachází společně s dalšími slovy patřícími do stejné kategorie. Takto například UNS rozpoznává slovo *cat* jako jasně odlišitelné od slova *mouse*. Přitom jej zároveň vnímá jako slovo, které má ke slovu *mouse* – ve srovnání s dalšími slovy – relativně nejbližší. Tzn. že obě slova pojímá jako (v něčem) sobě podobná. Tato dvě slova potom mají relativně nejbližší ke slovu *dog*. Společně tak tvoří skupinu slov označujících středně velké nebo drobnější živočichy. Jinak řečeno, UNS rozpoznává slovo *cat* jako označení středně velkého nebo drobnějšího živočicha, neboť jeho odpovídající aktivační vektor je na jedné straně podobný dalším aktivačním vektorům slov označujících jiné středně velké nebo menší živočichy a na straně druhé je odlišný od aktivačních vektorů slov označujících všechna ostatní slova. Na základě stejného principu pak UNS také dokáže rozpoznat, že slovo *cat* zároveň označuje nějakého živočicha, resp. nějakou živou bytost a že



Obrázek 156: Dendrogram hierarchické shlukové analýzy aktivačních vektorů, které byly vyvolány ve skryté vrstvě prezentací jednotlivých slov. (převzato z Elman, 2004)

patří mezi podstatná jména. V jednom aktivačním vektoru je tak zakódováno vše, co UNS ví o daném slově. Tyto její znalosti jsou přitom hierarchicky strukturovány - viz obrázek 157, na kterém je schematicky zachycen vnitřní reprezentační prostor Elmanovy UNS, kde jsou jednotlivá slova seskupena do několika různých množin a podmnožin odpovídajících výše zmiňovaným gramatickým a sémantickým kategoriím. Elman (2004) v této souvislosti také navrhuje dívat se na význam slova nikoli jako na nějakou informaci o jeho sémantice, syntaxi či výslovnosti, která je uložena na určitém konkrétním místě v pasivní paměťové struktuře, na které operují různé mentální procesy, ale jako na něco, co je důsledkem přímého kauzálního působení smyslové (jazykové i nejazykové) stimulace na vnitřní mentální stavy, ve kterém se odráží specifické sémantické, syntaktické a další charakteristiky daného slova použitého

v daném kontextu. Slovo v tomto pojetí tak není přímo nositelem svého významu, ale spíše jen jedním (přesto tím nejvýznamnějším) z několika vodítek k významu, který se druhá osoba pokouší sdělit. V podobném duchu Rumelhart (1979) tvrdí, že „porozumění [jazyku] by mělo být, stejně jako vnímání, připodobněno k Hebbovu (1949) paleontologovi, který používá svých přesvědčení a znalostí o dinosaurech ve spojení s vodítky poskytnutými dostupnými kosterními pozůstatky k vytváření plnohodnotného modelu originálu. V tomto případě by vyřčená slova a činnost mluvčího byla připodobněna k archeologovým vodítkům a *dinosaurus* k významu zprostředkovanému těmito vodítky.“¹

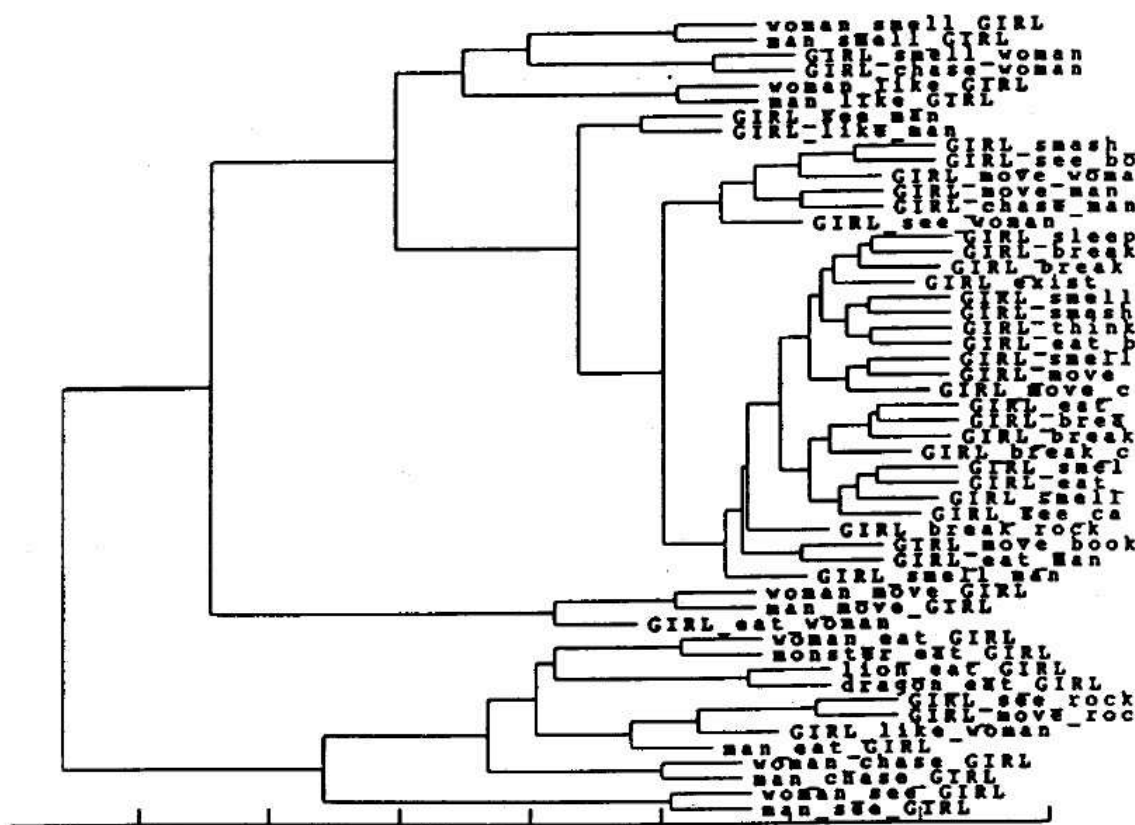


Obrázek 157: Schematické znázornění vnitřního reprezentačního prostoru Elmanovy UNS. Ve skutečnosti má její „mentální prostor“ 150 dimenzí. (převzato z Elman, 2004)

Dendogram zobrazený na obrázku 155 vznikl na základě hierarchické shlukové analýzy zprůměrovaných aktivačních vektorů vyvolaných jednotlivými slovy napříč všemi kontexty,

¹ „comprehension, like perception, should be likened to Hebb’s (1949) paleontologist, who uses his beliefs and knowledge about dinosaurs in conjunction with the clues provided by the bone fragments available to construct a full-fledged model of the original. In this case the words spoken and the actions taken by the speaker are likened to the clues of the paleontologist, and the dinosaur, to the meaning conveyed through these clues.“

ve kterých se dané slovo mohlo vyskytnout. Dendogram tak zakrývá jednu důležitou skutečnost, totiž že Elmanova UNS vykazovala také senzitivitu vůči kontextu, ve kterém se slovo nacházelo. Jinými slovy, jedno a to samé slovo vyskytující se v různém kontextu UNS ve své skryté vrstvě reprezentovala odlišným (přesto – ve srovnání s dalšími slovy – velice podobným) způsobem. Na obrázku 158 je takto zobrazen dendogram zachycující strukturu podobnosti aktivačních vektorů odpovídajících slovu *girl* vyskytujícímu se v různých kontextech. Z dendogramu je dobře patrné, že UNS rozlišuje například mezi slovy *girl*, která hrají roli gramatického subjektu nebo objektu. Jedná se de facto o implementaci v lingvistice velice dobře známého rozdílu mezi typem (*type*) a znakem (*token*), kde typ představuje abstraktní reprezentaci slova, která zahrnuje všechny důležité informace o daném slově týkající se jeho sémantiky, výslovnosti a jeho syntaktických charakteristik, a znak je konkrétní realizací takového typu v nějaké jazykové výpovědi. Například ve větě *Po startu se Hamiltonova formule ocitla těsně vedle formule Fernanda Alonsa* první a druhé použití slova

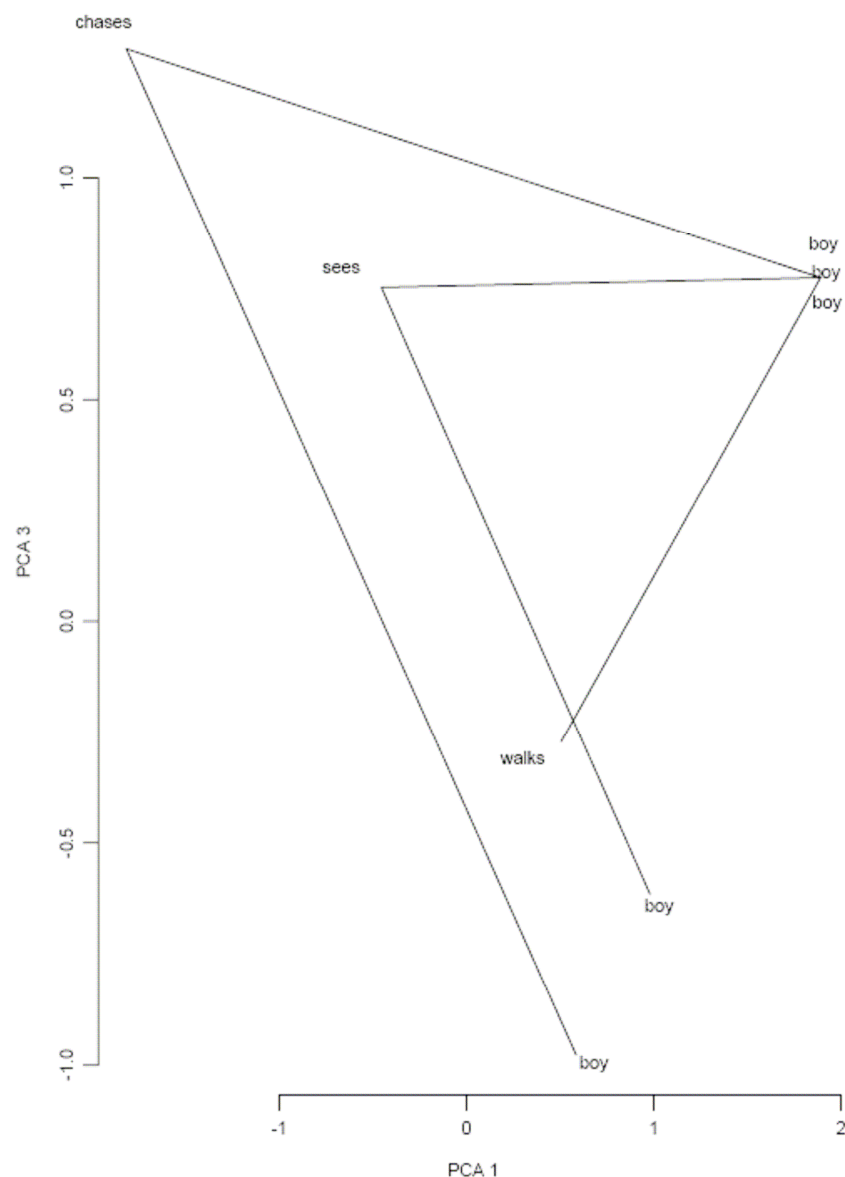


Obrázek 158: Dendogram zachycující strukturu podobnosti aktivačních vektorů ve skryté vrstvě vyvolaných slovem *GIRL* nacházejícím se v různých kontextech. (převzato z Elman, 1990)

formule odkazuje ke dvěma rozdílným monopostům, které však patří do stejné kategorie objektů. V UNS je typ implicitně reprezentován tak, že znaky jednoho typu se v jejím reprezentačním prostoru nacházejí ve stejné oblasti, a že tvoří jasně vydělitelnou skupinu aktivačních vektorů, které si jsou navzájem mnohem podobnější než s jakýmkoli jinými aktivačními vektory odpovídajícími ne-znakům daného typu. Takto například slovo *girl* nacházející se v různých kontextech bude mít ve skryté vrstvě vždy trochu odlišný aktivační vektor, přesto si budou jednotlivé aktivační vektory natolik podobné, že jimi definované pozice jednotlivých znaků (typu GIRL) se budou vyskytovat v jedné omezené oblasti vnitřního reprezentačního prostoru UNS.

Elman (1993) ve své další práci prokázal, že jednoduché rekurentní sítě (anticipující svůj příští vstup) si po náležitém tréninku dokáží osvojit i složitější gramatické jevy jako je například shoda podmětu s přísudkem, a to i ve větách, které obsahují vztažnou větu. Nová simulace měla podobné parametry jako ta předchozí, jen s tím zásadním rozdílem, že v tréninkovém prostředí UNS se nacházel mnohem realističtější vzorek anglického jazyka, tzn. že v něm bylo možné „narazit“ na substantiva jak v jednotném tak i v množném čísle, na slovesa v základním tvaru nebo v tvaru odpovídajícím 3. osobě jednotného čísla a na složitější věty, včetně těch obsahujících v sobě nějakou vztažnou větu. Při zpracovávání tohoto typu vět si UNS musí umět zapamatovat, že podmět dané věty je v singuláru nebo v plurálu, neboť jen tak dokáže předpovědět sloveso ve správném tvaru (tj. s koncovkou –s nebo bez ní, alespoň tedy v případě anglického jazyka). Tento úkol je o to těžší, že mezi podmětem a odpovídajícím slovesem se nachází vztažná věta. Stejně jako v předchozí simulaci, ani nyní nebyla UNS schopna přesně predikovat jedno konkrétní slovo, neboť možných kandidátů bylo vždy větší množství. Dokázala však anticipovat množinu gramaticky (a sémanticky) přijatelných slov, které by mohly následovat. Takže když například podmětem věty bylo nějaké substantivum v jednotném čísle, UNS dokázala správně předpovědět, že sloveso následující za vztažnou větou musí mít tvar odpovídající 3. osobě jednotného čísla (tedy že musí mít koncovku –s).

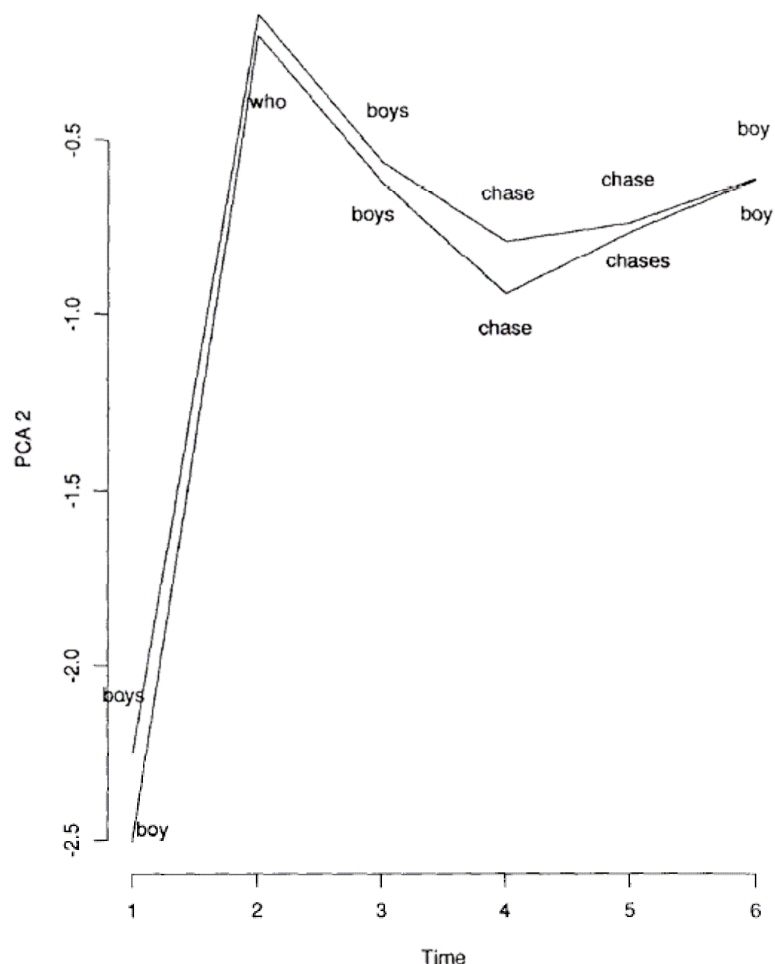
K analýze vnitřního reprezentačního prostoru UNS Elman tentokrát použil statistickou metodu analýzy hlavní komponent (dále AHK). S její pomocí pak zkoumal vývojové trajektorie neuronové sítě v jejím stavovém prostoru v průběhu zpracovávání různých druhů vět. Na obrázku 159 jsou takto zobrazeny tři různé vývojové trajektorie, které zachycují dynamiku (vnitřního) chování UNS při zpracovávání tří různých vět (*boy chases boy*, *boy sees boy* a *boy walks*), které sestávají ze stejných substantiv, ale z odlišných sloves, která se liší v gramatické kategorii tranzitivity: Zatímco u jednoho slovesa (*chases*) je přímý objekt



Obrázek 159: Vývojové trajektorie opisované UNS v jejím stavovém (vnitřním reprezentačním) prostoru během zpracovávání tří různých vět („boy chases boy”, „boy sees boy” a „boy walks”), jejichž slovesa se liší v míře tranzitivity; více viz hlavní text. (převzato z Elman, 1993)

nezbytně nutný, druhé sloveso (*sees*) ho mít může (ale nemusí) a třetí sloveso (*walks*) jeho použití vůbec neumožňuje. Tento rozdíl v míře tranzitivity slovesa se v chování UNS projevuje tím, jaké oblasti stavového prostoru UNS přitahují její vývojové trajektorie při zpracovávání těchto sloves. Z obrázku je dobře patrné, že UNS tranzitivitu slovesa kóduje podél osy, která prochází napříč první a třetí hlavní komponentou: Čím je sloveso tranzitivnější, tím více se nachází v levém horním rohu; čím je naopak méně tranzitivní, tím více se nachází v pravém dolním rohu. Podobným způsobem UNS kóduje také rozdíl mezi substantivy *boy* podle toho, zda jsou použita jako gramatický podmět nebo jako gramatický

předmět: Zatímco v prvním případě je vývojová trajektorie UNS přitahována do horní poloviny zde zobrazeného řezu stavovým prostorem, v druhém případě je vývojová trajektorie přitahována směrem k jeho dolní polovině. Gramatická struktura věty jako celku se tak zrcadlí ve tvaru vývojové trajektorie, kterou ve svém stavovém prostoru UNS opisuje při zpracovávání dané věty. Na jazyk, resp. na jeho zpracovávání (ve smyslu produkce či porozumění) je tak možné se dívat jako druh dynamického systému, který lze charakterizovat v termínech teorie dynamických systémů, tj. prostřednictvím takových pojmů jako jsou stavový prostor, vývojová trajektorie, atraktor, repelér, basin of attraction atd. (více viz Elman, 1995). Na obrázku 160 jsou zobrazeny vývojové trajektorie UNS napříč jednou z dimenzí jejího stavového prostoru při zpracovávání dvou téměř identických vět (*boy who boys chase chases boy* a *boys who boys chase chase boy*), které se skládají ze stejných slov, obsahují stejnou vztahnou větu, ale liší se v čísle podmětu hlavní věty a tím pádem také ve



Obrázek 160: Vývojové trajektorie UNS napříč jednou z dimenzí jejího stavového prostoru, ve které je zakódován rozdíl mezi podmětem v jednotném a množném čísle ve větách „boy who boys chase chases boy“ a „boys who boys chase chase boy“. (převzato z Elman, 1993)

tvary odpovídajícího slovesa. Jak již zde bylo uvedeno, k tomu, aby UNS dokázala predikovat gramaticky správnou skupinu sloves, musí si umět zapamatovat, v jakém čísle je podmět hlavní věty. Z obrázku je tato paměťová stopa dobře patrná z mírně odlišné vývojové trajektorie při zpracovávání těchto dvou různých vět. Tyto drobné rozdíly způsobují, že v jednom případě vývojová trajektorie skončí v oblasti stavového prostoru, která odpovídá slovesu v základním tvaru, a v druhém případě v oblasti odpovídající tvaru slovesa ve 3. osobě jednotného čísla. Tato analýza vnitřních „mentálních“ reprezentací UNS tak alespoň naznačuje, jak je možné, že si UNS dokáže poradit se shodou podmětu s přísudkem i ve větách, které v sobě obsahují nějakou vztažnou větu – navzdory tvrzení některých kognitivních vědců (Fodor, Pylyshyn, 1988), že něco takového stojí naprosto mimo dosah schopností UNS (tedy alespoň v tom případě, že neimplementují nějaký klasický symbolový systém; viz také oddíl „1.2.3 Výpočetní teorie mysli,“).

Další, z vývojového hlediska velice zajímavou charakteristikou této simulace bylo to, že aby se UNS dokázala naučit předpovídat gramaticky přijatelné skupiny slov i ve větách s vloženými vztažnými větami, musela si nejdříve osvojit základní gramatickou (a sémantickou) strukturu svého jazyka na jednodušších dvoj- nebo trojslovních větách. Teprve poté, co si určitým specifickým způsobem uspořádala svůj vnitřní reprezentační prostor (tak aby odrážel gramatické a sémantické zákonitosti jí prezentovaného jazyka), mohla přistoupit k osvojování znalostí nutných ke zvládnutí shody podmětu s přísudkem ve složitějších větách s vztažnou větou. Kdykoli Elman vložil tento složitější typ vět do tréninkového prostředí UNS ihned od začátku jejího tréninku, UNS se potřebným znalostem nikdy nedokázala naučit. Když však v tréninkovém prostředí byly zpočátku jen jednodušší věty a teprve poté se do něj postupně přidávalo stále větší množství složitějších vět, UNS se nakonec dokázala úspěšně naučit anticipovat gramaticky vhodná slovesa i v případě složitějších vět obsahujících v sobě nějakou vztažnou větu (tzv. *inkrementální učení*, *incremental learning*). Elmanův model se tak jeví implementovat (aniž by to přitom bylo nějak předem plánováno – je to zkrátka jen vedlejší efekt toho, jakým způsobem UNS zpracovává informace) jednu z výrazných charakteristik intuitivního rodičovství, kdy dospělé osoby mluví na malé děti (v rámci tzv. řeči zaměřené na dítě) v gramaticky jednodušších větách (viz oddíl „Protosociální chování dítěte, intuitivní rodičovství a „maminkovština““). Elmanova simulace pak naznačuje, jaké by mohlo být racionále tohoto aspektu intuitivního rodičovství. Později Elman dosáhl stejného efektu prostřednictvím omezení vnitřních výpočetních zdrojů UNS: V každém třetím nebo čtvrtém kroku Elman vymazal obsah kontextových jednotek, čím značně zredukoval kapacitu její paměti a znemožnil jí tím zpracovávat věty, které měly více

než čtyři slova. Důsledek byl tedy stejný jako v případě, kdy se do tréninkového prostředí zpočátku vkládaly jen kratší věty. Elman poté postupně snižoval množství šumu omezujícího kapacitu paměti UNS, tak aby se tato dokázala učit také vztahům a závislostem mezi vzdálenějšími větnými členy. Stejně jako při omezení vnějšího podnětového prostředí tak i při omezení vnitřních výpočetních zdrojů se UNS nakonec dokázala úspěšně naučit anticipovat gramaticky přípustná slova - a to jak v případě známých vět vyskytujících se v tréninkovém prostředí, tak i v případě vět, které UNS měla možnost vidět poprvé.¹ Elman tímto demonstroval, že omezená kapacita pracovní paměti u malých dětí nemusí být pouze „nechtěným“ a „trpěným“ důsledkem nezralosti jejich CNS, ale že může také plnit určitou adaptivní funkci při osvojování některých kognitivních funkcí. Jinými slovy, Elman demonstroval, že při osvojování kognitivních funkcí - tak jako v mnoha jiných oblastech života - platí okřídlené rčení, že „méně je někdy více“.

3.2.3.5.5 Empirický výzkum osvojování gramatiky

To, že také malé děti dokáží určitým způsobem využívat statistických zákonitostí ve výskytu jednotlivých slov v jazykovém kontextu k vyvozování gramatických pravidel jazyka, demonstrovaly ve svém výzkumu Gomezová a Gerkenová (1999), které nejdříve vytvořily jednoduchý umělý jazyk s vlastní gramatikou, poté dvanáctiměsíčním dětem prezentovaly věty vygenerované na základě této gramatiky a nakonec sledovaly, jak tyto děti budou (ve smyslu preference) reagovat na prezentaci nových vět, které byly vytvořeny buď v souladu, nebo naopak v rozporu s gramatickými pravidly umělého jazyka. Zjistily přitom, že děti měly statisticky významnou tendenci preferovat ty věty, které se řídily stejnými gramatickými pravidly, podle kterých byly vytvořeny věty prezentované jim během familiarizační fáze experimentu.

To spolu s řadou dalších výzkumů naznačuje, že děti mohou vykazovat citlivost vůči různým gramatickým jevům ještě dlouho předtím než začnou sami aktivně používat svých znalostí gramatiky k produkci vlastních (gramaticky složitějších) vět. Takto například některé práce z Jusczykovy laboratoře ukazují, že zhruba devatenáctiměsíční děti z anglofonního prostředí preferují gramaticky správné věty, ve kterých je dodrženo gramatické pravidlo

¹ Novými větami se zde pochopitelně myslí nové a dosud nepotkané kombinace již známých slov, neboť jen v jejich případě UNS disponuje potřebnými znalostmi o jejich gramatických a sémantických vlastnostech, které může využít při generalizaci na nové věty.

shody podmětu s přísudkem¹ (Soderstrom, Wexler, Jusczyk, 2002) a gramatické pravidlo pro tvorbu průběhového času prostřednictvím pomocného slovesa *is* (nebo *am* či *was*) a koncovky *-ing* (Santelmann, Jusczyk, 1998; Tincoff, Santelmann, Jusczyk, 2000). Jiné studie pak za využití experimentální techniky intermodální zrakové preference² prokázaly, že již sedmnáctiměsíční (ale zcela určitě 29 měsíců staré) děti z anglofonního prostředí dokáží využívat znalosti pravidla slovosledu podmět-přísudek-předmět k určení toho, co je ve větě podmětem a co předmětem (a v souladu s tím pak zaměřit svou pozornost na odpovídající scénu).

Ukazuje se, že při osvojování gramatiky se uplatňují také prozodická vodítka, na základě kterých již devítiměsíční děti dokáží v řečovém projevu identifikovat jednotlivé syntaktické složky věty jako je jmenná a slovesná fráze, resp. podmětová a přísudková část věty. Tuto svou schopnost děti v experimentálních podmínkách projevují například tím, že zřetelně preferují ty posloupnosti slov, které v řečovém projevu tvoří nějakou syntaktickou složku věty, a jako takové tedy vykazují také tu správnou „prozodickou signaturu“ (Soderstrom, Jusczyk, Kemler Nelson, 2000).

3.2.4 Přínos konekcionistického modelování pro výzkum (nejen) jazykového vývoje

Práce s umělými neuronovými sítěmi v rámci (konekcionistického) modelování kognitivního vývoje přinesla (a stále přináší) řadu zajímavých výsledků, které vrhají zcela nové světlo na některé z tradičních otázek, na které se vývojová psychologie, resp. vývojová psycholinguistika snaží (více či méně úspěšně) najít odpověď. Několik následujících oddílů je proto věnováno tomu, jaký je (nebo jaký by mohl být) přínos konekcionistického modelování pro porozumění některým základním vývojově-psychologickým (a vývojově-psycholinguistickým) problémům jako je například vztah dědičnosti a prostředí při osvojování jazyka, povaha samotného kognitivního vývoje či osvojovaných poznatků atp.

¹ Konkrétně se jednalo o koncovku slovesa ve 3. osobě jednotného čísla *-s*.

² Tato experimentální technika se většinou používá ke zjišťování porozumění řeči. Podstata této metody spočívá v tom, že dítěti jsou nejdříve současně prezentovány dvě různé scény (v podobě obrázku nebo nějaké videosekvence) a poté je slovně instruováno, aby se dívalo na jednu konkrétní scénu. Na základě porovnání délky trvání pohledu, který dítě věnuje pozorování jednotlivých scén, pak lze usuzovat na to, zda dítě dokázalo porozumět slovní instrukci, případně zda projevuje citlivost vůči určitému gramatickému jevu (například větnému slovosledu), který je ve slovní instrukci zakódován (Smolík, 2006).

3.2.4.1 Nativismus vs. empirismus

Dějiny psycholingvistiky se od jejích samotných počátků jako červená nit vine tradiční otázka všech věd zabývajících se v té či oné podobě lidským chováním. Tato otázka se týká toho, jakou roli hrají dědičnost (*nature*) a prostředí (*nurture*) v procesu osvojování různých schopností. Na tuto otázku existují v zásadě tři možné odpovědi: 1) První typ odpovědi předpokládá, že schopnosti a s nimi spojené poznatky jsou člověku (v té či oné podobě) vrozené, tzn. že člověk se s nimi již rodí a nemusí se jim v pravém slova smyslu učit – jejich osvojování má spíše podobu zrání nebo jakési inicializace. 2) Druhý typ odpovědi naopak předpokládá, že člověk přicházející na svět je nepopsaná deska a že vše, co člověk umí a zná, je výsledkem procesu učení na základě zkušeností s vnějším prostředím. 3) Třetí typ odpovědi pak nabízí jistý kompromis mezi předchozími dvěmi možnostmi, když navrhuje, že schopnosti a znalosti jsou výsledkem interakce mezi vrozenými dispozicemi a vnějším prostředím, které vrozené schopnosti v různé míře modifikuje a kalibruje, tak aby tyto byly v souladu s lokálními podmínkami. Tyto tři základní typy odpovědí na otázku po vztahu dědičnosti a prostředí lze v jisté rudimentární podobě nalézt již v novověké filozofii 17. a 18. století, která v mnohém vymezila „duševní obzory“ současného západního myšlení v oblasti společenských věd.¹ Typickým představitelem prvního přístupu byl *René Descartes* (1596-1650), který předpokládal, že člověk se již rodí vybaven řadou vrozených idejí. Existenci jakýchkoli vrozených idejí naopak tvrdě odmítal anglický filozof *John Locke* (1632-1704), který tvrdil, že veškeré obsahy lidské mysli mají v posledku svůj zdroj ve vnější zkušenosti. Odtud také jeho slavný výrok, že nic není v rozumu, co dříve nebylo ve smyslech, a jeho neméně slavné pojetí lidské mysli jako prázdné voskové tabulky (*tabula rasa*), která je teprve v průběhu života člověka zaplňována různými počitky a dojmy. Konečně třetí přístup měl svého zastánce v osobě filozofa *Gottfrieda Wilhelma Leibnize* (1646-1716), podle kterého je sice pravda, že v mysli se nikdy nenachází nic, co by předtím nebylo ve smyslech, ovšem s důležitou výjimkou mysli samotné. V otázce role vrozených předpokladů a prostředí při osvojování jazyka lze v dějinách psycholingvistiky nalézt obdobné názorové proudy jako v novověké filozofii a to dokonce i v podobném chronologickém pořadí.²

¹ Tento vliv se projevuje například ve stále přežívajícím „strašidle“ „karteziánského divadla“ (Dennett, 1993), na které můžeme narazit v nejedné moderní teorii některého z kognitivních procesů (viz také s. 54). Jednou z jeho nejznámějších současných „inkarnací“ jsou exekutivní (řídící) funkce, které ze svého „velitelského stanoviště“ v prefrontálním kortexu „dirigují“ činnost ostatních částí mozku.

² Což není až zas tolik překvapující, vezmeme-li v úvahu lidskému myšlení zcela vlastní (vrozenou?) tendenci k „odezdikezdismu“, kterou německý filozof *Georg Wilhelm Friedrich Hegel* (1770-1831) mnohem

Pomineme-li průkopnickou, ale ne příliš vlivnou Wundtovu práci *Die Sprache* (1900), ve které její autor upozorňoval na to, že při studiu jazykových schopností člověka by se měla stejná pozornost jako samotnému jazyku věnovat také lidské mysli, která jazyk používá, první psychologickou školou, která zásadním způsobem ovlivnila způsob uvažování o povaze procesu osvojování jazyka, byl behaviorismus. Ten (ve své klasické podobě) při vysvětlování lidského chování zcela rezignuje na snahu poznat vnitřní obsahy a procesy lidského myšlení a soustředí se pouze na studium objektivně pozorovatelných vztahů mezi souborem podnětů a jimi vyvolaných reakcí (Hoskovec, Nakonečný, Sedláková, 2002). Veškeré lidské chování tak behavioristé chápou jako reakce na podněty z prostředí. Tuto základní myšlenku behaviorismu zachycuje schéma $R = f(S)$, kde chování (R jako *response*) je funkcí (f) podnětů z prostředí (S jako *stimulus*). To, jakým konkrétním způsobem dochází k asociování jednotlivých behaviorálních reakcí a podnětů, objasňuje behavioristická teorie učení, která popisuje několik (vše)mocných učících mechanismů (*klasického a operantní podmiňování*, nazývaného také *učení pokusem a omylem*, resp. *učení postupnou aproximací*), které stojí za vznikem všech vzorců lidského chování. Tzn. že podle behavioristů je veškeré lidské chování naučené a k jeho vysvětlení není potřeba předpokládat existenci žádných vrozených dispozic. Plně symptomatické je v tomto ohledu smělé prohlášení zakladatele klasického behaviorismu Johna B. Watsona: „Dejte mi tucet zdravých, dobře stavěných dětí a můj vlastní svět, ve kterém bych je mohl vychovávat, a já vám ručím za to, že z kteréhokoli z nich, náhodně vybraného, mohu vychovat jakéhokoli odborníka – lékaře, právníka, umělce, obchodníka, velitele a samozřejmě i žebráka nebo zloděje, nezávisle na jejich talentu, sklonech, schopnostech, povolání a rase jejich předků.“¹ (Watson, 1930, s. 104) Ve Watsonových očích je „lidská mysl“ neomezeně tvárným materiálem, se kterým lze (prostřednictvím posilňování určitých návyků) libovolně manipulovat a učit ho se chovat požadovaným způsobem. Stejně výkladové schéma, které behavioristé použili k vysvětlení jednodušších forem chování, se snažili aplikovat rovněž na jazykové chování a na proces osvojování jazyka. Vyvrcholením snah v tomto směru byla Skinnerova kniha *Verbal behaviour* (Verbální chování) z roku 1957.

Ve stejném roce americký psycholingvista Noam Chomsky vydal knihu *Syntactic structures* (Syntaktické struktury), ve které se podrobně věnoval pojmu gramatických pravidel. Stejný autor (Chomsky, 1959) pak v recenzi ke Skinnerově knize *Verbal behaviour*

sofistikovaněji popsal jako *dialektický proces* myšlení, který se odehrává v krocích *teze* (tvrzení) → *antiteze* (opak předchozího tvrzení) → *syntéza* (tvrzení zachovávající některé z prvků dvou protirečících si tvrzení).

¹ „Give me a dozen healthy infants, well-formed, and my own specified world to bring them up in and I'll guarantee to take any one at random and train him to become any type of specialist I might select – doctor, lawyer, artist, merchant-chief and, yes, even beggar-man and thief, regardless of his talents, penchants, tendencies, abilities, vocations, and race of his ancestors.“

zpochybnil víru behavioristů, že by bylo možné vysvětlit produktivitu a systematickosti (*systematicity*)¹ jazyka na základě asociačních řetězců jazykových podnětů a reakcí vznikajících prostřednictvím učicího mechanismu klasického nebo operantního podmiňování. Podle Chomskyho lze vysvětlit schopnost člověka generovat téměř neomezené množství gramaticky správných vět jedině tak, že budeme předpokládat, že člověk ve své mysli disponuje určitými abstraktními (gramatickými) pravidly, která specifikují, jak se mohou slova kombinovat do frází a do vět podle jejich příslušnosti k jednotlivým slovním druhům. Soubor těchto pravidel, která si člověk nosí ve své mysli, Chomsky nazývá **generativní (a transformační)² gramatikou**. Používání jazyka tak podle Chomskyho předpokládá existenci určitých mentálních reprezentací. Tímto svým názorem Chomsky v psycholingvistice způsobil malou (kognitivní) revoluci, neboť do ní (znovu)zavedl pojmy mentalistického diskurzu. Tento pohled na jazyk byl přitom v souladu s právě tehdy se vynořující počítačovou metaforou lidské mysli, která lidskou mysl chápala jako výpočet ve smyslu manipulace a transformace explicitních symbolických struktur, resp. jako druh kombinatorického výpočetního systému, který se skládá ze souboru systematicky kombinovatelných symbolických reprezentací a ze souboru procedur, které kromě běžné manipulace umožňují také dekompozici komplexních symbolických struktur na symbolové atomy či naopak jejich kombinaci do složitějších symbolových molekul (Fodor, Pylyshyn, 1988).

Používání jazyka (v Chomskyho terminologii označované jako *performance*) je tedy podle Chomskyho založeno na znalosti komplexního systému gramatických pravidel (v Chomskyho terminologii označovaného jako *competence*),³ která mluvčímu umožňuje rozpoznávat a také tvořit gramaticky správné věty. Zůstává však otázka, jak si člověk tento systém pravidel osvojuje. Vzhledem ke složitosti gramatických pravidel spojených s používáním jazyka je až s podivem, jakou rychlostí si malé děti dokáží tato pravidla v prvních třech čtyřech letech

¹ Systematickosti jazyka zde označuje tu skutečnost, že generování vět není náhodné, ale systematické, tj. že se řídí určitým souborem (v daném jazyce) obecně platných pravidel, která specifikují, jak lze s pomocí jednoduchých stavebních prvků skládat složitější struktury a naopak, jak lze složitější struktury rozložit na jejich základní stavební prvky.

² Pravidla transformační gramatiky se vztahují k návodům, jak převádět **hloubkovou syntaktickou strukturu** (*deep structure*) na **povrchovou syntaktickou strukturu** (*surface structure*). Podle Sternberga (2002, s. 335) Chomsky „pojem hloubková struktura užil... pro základní syntaktické struktury, které prostřednictvím transformačních pravidel propojují [věty s různými frázovými strukturami]. Pojem povrchová struktura [pak] užil pro rozmanité frázové struktury, které mohou být výsledkem těchto transformací.“ Rozlišení mezi hloubkovou a povrchovou strukturou tak umožňuje určit vztah mezi různými frázovými strukturami a „postihovat [tu] skutečnost, že věty s podobnou povrchovou strukturou mohou mít vzájemně velmi odlišný význam [a vice versa].“ (ibid., s. 334)

³ Chomskyho termínům *competence* a *performance* do značné míry odpovídají de Saussurovy pojmy *langue* a *parole*, z nichž ten první označuje soubor abstraktních gramatických pravidel a ten druhý pak běžnou mluvenou řeč, která vzniká realizací abstraktních gramatických pravidel v nějaké konkrétní jazykové promluvě (Budil, 1998, s. 142)

života úspěšně osvojit. To, co je na procesu osvojování jazyka také nápadné a pozoruhodné zároveň, je to, že děti úspěšně generalizují za rámec jazykového „materiálu“, se kterým se během svého života stačily potkat. Tzn. že děti dokáží rozumět větám (a také vytvářet věty), které nikdy předtím neslyšely, a to včetně těch, jejichž gramatická struktura přesně neodpovídá gramatické struktuře vět, se kterými měly možnost se běžně potkávat. To jen potvrzuje tu skutečnost, že používání jazyka není založeno na nějakém jednoduchém napodobování nebo na mechanickém zapamatovávání si přípustných kombinací slov do frází a vět, ale že je za ním skrytá znalost abstraktních gramatických pravidel. Problém s jejich osvojováním je v tom, že jazykový vstup je podle všeho mnohem chudší než jazykový výstup (*argument chudé stimulace, poverty of the stimulus argument*). Důsledkem je to, že s jazykovým vstupem, který je dítěti dostupný, je kompatibilních až příliš mnoho různých souborů gramatických pravidel. Dítě, které se snaží osvojit gramatická pravidla svého mateřského jazyka, tak stojí před problémem, jak určit, který z těch mnoha možných souborů gramatických pravidel je ten správný. To podle Chomskyho znamená, že gramatice svého mateřského jazyka se dítě nemůže v žádném případě naučit pouze na základě svých zkušeností s jazykovým prostředím, ve kterém vyrůstá. Jinak řečeno, podle Chomskyho osvojování jazyka nemůže být záležitostí čistě jenom učení. Známý je v této souvislosti Goldův (1967) formální důkaz, podle kterého není možné se naučit gramatice žádného přirozeného jazyka pouze na základě tzv. pozitivních příkladů, tj. pouze na základě příkladů gramaticky správných vět a bez jakýchkoli negativních příkladů spočívajících v explicitním konstatování (rodiči nebo jinými dospělými mluvčími), že určitý typ větné konstrukce není v daném jazyce gramaticky přípustný. Vzhledem k tomu, že rodiče jen málokdy své děti explicitně opravují, když vytvoří nějakou větu, která není gramaticky zcela správně, existují v zásadě pouze dvě možná řešení této situace: Buď existuje nějaká méně nápadná forma negativních příkladů, kterou děti dokáží využívat, nebo děti disponují určitými vrozenými znalostmi o některých aspektech jazyka, které pak mohou využívat ke zmenšení množiny možných kandidátů na „pozici“ souboru gramatických pravidel jejich mateřského jazyka. Podle Chomskyho tím, co dítěti umožňuje si jazyk osvojit, jsou vrozené predispozice. Chomsky (1975) v této souvislosti hovoří o tzv. **univerzální gramatice** (*Universal Grammar*) - souboru vrozených jazykových principů, bez jejichž existence by nebylo možné vysvětlit ty vlastnosti jazyka, u kterých nelze předpokládat, že by se je dítě mohlo samo naučit na základě pozitivních příkladů z jazykového prostředí. Chomskyho (1975) oblíbeným příkladem vrozeného jazykového principu je princip strukturní závislosti (*structure-dependence*), který se týká vrozené tendence dítěte automaticky upřednostňovat ty hypotézy ohledně

gramatických vlastností jazyka, které jsou v souladu s předpokladem, že jazyk tvoří komplexní hierarchickou strukturu a nikoli pouze lineární sled slov, přestože dostupné jazykové vstupy jsou často konzistentní s oběma typy hypotéz (tj. jak s principem strukturní závislosti tak i s principem strukturní nezávislosti). Jako doklad vrozenosti tohoto principu Chomsky uvádí způsob, jakým se děti učí tvorbě otázek: Přestože se v jazykovém vstupu malého dítěte téměř nevyskytují žádné otázky typu (1) *Is the man who is smoking crazy?*, děti se vždy naučí tvořit tento typ otázek správným způsobem, takže nikdy během svého jazykového vývoje nedělají tu chybu, že by produkovaly otázky typu (2) *Is the man who smoking is crazy?* Podle Chomskyho to znamená, že děti se tvorbě otázek učí na základě jednoduchých oznamovacích vět jako například (3) *The man is here* → (4) *Is the man here?*. Na základě těchto jednoduchých vět však dítě může vyvodit dvě různá pravidla: První pravidlo je v souladu s předpokladem strukturní nezávislosti a má podobu jednoduchého návodu *Postupuj slovo za slovem a když narazíš na „is“, dej ho na začátek věty*. Druhé (a správné) pravidlo naopak vychází z předpokladu strukturní závislosti a dítě se na jeho základě řídí „radou“ *Postupuj slovo za slovem a na začátek věty dej jen takové „is“, které se nachází za první jmennou frází*. Skutečnost, že děti nevytvářejí otázky typu (2), svědčí o tom, že z těchto dvou pravidel děti upřednostňují druhé pravidlo, a to přesto, že se jedná o mnohem složitější pravidlo a že jazykové vstupy jsou stejně kompatibilní i s druhým pravidlem. To podle Chomskyho znamená, že „*jediným rozumným závěrem je, že [univerzální gramatika] obsahuje princip, podle kterého všechna taková pravidla musí vykazovat strukturní závislost.*“¹ (Chomsky, 1975)

Chomskyho hypotézu o vrozenosti některých jazykových principů se zdají podporovat výsledky řady empirických výzkumů a pozorování. Vedle zde již zmiňované rychlosti, s níž si děti dokáží jazyk osvojit, nízkého věku, kdy k tomuto procesu dochází, a schopnosti dětí generalizovat za rámec (omezeného) jazykového vstupu to jsou například stejné fáze v jazykovém vývoji pozorované napříč různými jazykovými společenstvími, a to včetně mluvních znakového jazyka; jsou to stovky **jazykových univerzálií** - vlastností, které lze nalézt téměř ve všech jazycích; takto lze například v každém jazyce narazit na slovní a větné druhy, na pomocná slovesa, na flexi atp. Dále je to existence vysoce specializovaných oblastí mozkové kůry, které jsou určeny výhradně k reprezentaci řečových funkcí, které jsou u většiny lidí (praváků i nepraváků) lokalizovány v levé mozkové hemisféře. Jedním z nejsilnějších dokladů pro vrozenost jazyka je *kreolizace* – proces, kdy děti rodičů, kteří

¹ „...the only reasonable conclusion is that UG contains the principle that all such rules must be structure-dependent.”

mluví nějakou formou *pidžin* jazyka, který nemá téměř žádnou gramatiku (tj. je bez gramatických morfémů, bez předložek, bez členů, obsahuje pouze „telegrafické věty“ atd.), dokáží z tohoto neúplného jazyka učinit jazyk s vlastní plnohodnotnou gramatikou, tzv. *kreolštinu*. Lingvista *Derek Bickerton* (1981) ve své knize *Roots of language* (Kořeny jazyka) popsal, jak tento proces kreolizace proběhl někdy na přelomu 19. a 20. století na Havajských ostrovech, kde se sešla pestrá populace lidí pocházejících z různých koutů světa (z Číny, Japonska, Portugalska či Filipín), kteří si ke společné komunikaci a sdělování základních informací vytvořili jednoduchý pidžin jazyk, který byl směsicí výrazů z jejich původních jazyků postrádající jakoukoli složitější gramatiku. Tu do jazyka dodala až následující generace, tzn. děti přistěhovalců, kteří pidžin jazyku byly vystaveny již od svého narození. Proces kreolizace byl pozorován také na konci 70. let 20. století v Nikaragui v tamější nově otevřené škole pro hluchoněmé, kteří si ke společné komunikaci nejdříve vytvořili znakovou obdobu pidžin jazyka (tzv. *Lenguaje de Signos Nicaragüense*, LSN) - ten obsahoval pouze jednoduché znakové výrazy bez možnosti vyjadřovat jakékoli složitější myšlenky, zápletky nebo příběhy. Avšak nová generace hluchoněmých dětí, která byla od nejútlejšího věku vystavena tomuto znakovému pidžinu, ho dokázala přetvořit v plnohodnotný znakový jazyk (*kreolštinu*) se složitou gramatikou umožňující rozlišit kdo udělal co a komu (tzv. *Idioma de Signos Nicaragüense*, ISN). To naznačuje, že proces osvojování jazyka není pouze nějakou pasivní recepcí již existující struktury, ale že se jedná také o aktivní (re)konstrukci jazyka, jehož útržky dítě okolo sebe může slyšet. *Steven Pinker* (1994) v této souvislosti hovoří přímo o **jazykovém instinktu** (*language instinct*), čímž se snaží zdůraznit tu skutečnost, že „jazyk je něco člověku tak vlastního, jako třeba uchopování věcí: tak jako úchop je dán vrozenou schopností stavět palec proti ostatním prstům, je i jazyk z podstatné části dán něčím vrozeným, instinktivním“. Jazyk tedy nepřichází k člověku „zvenčí“; je z jisté podstatné části dán tím, co má „uvnitř“.“ (*Peregrin*, 2003, s. 53)¹ Podobně *Chomsky* - v souladu s *Goldovým* argumentem o nemožnosti získat dokonalý jazykový výstup z nedokonalého jazykového vstupu - tvrdí, že jazyk je něco, co se člověk nemůže nikdy v pravém slova smyslu naučit, a

¹ Tento jazykový instinkt je přítom u dětí aktivní pouze po určitou omezenou dobu (zhruba do 5. roku věku) a „spouští“ se pouze za přítomnosti jazykových podnětů v prostředí - jak to dokládají příklady dětí, kterým se v kritickém období z různých důvodů (například proto, že vyrůstaly v naprosté sociální izolaci) nedostalo dostatečného množství jazykových podnětů, a kterým se proto již nikdy nepodařilo si osvojit jazyk ve vší jeho (gramatické) komplexitě. Dobře známý je v této souvislosti případ indických dívek *Kamaly* a *Amaly*, které byly vychovávány vlčicí, nebo případ dívky *Genie*, která od narození až do svých 11 let žila v naprosté izolaci v jedné místnosti připoutaná k židli. V obou dvou případech se dívky (přes veškeré úsilí a snahu jejich učitelů) nedokázaly naučit vytvářet věty se složitější gramatickou strukturou. Důležitost kritického období pro osvojení jazyka demonstroval ve svém výzkumu také *Meier* (1991), když prokázal, že existuje souvislost mezi dobou, kdy se hluchoněmý člověk začne učit znakové řeči, a úrovní, na jaké se mu podaří si osvojit různé gramatické nuance znakového jazyka po 30 a více letech jeho používání.

že se jedná spíše o druh zrání nějakého „mentálního orgánu“: „*O jazykové dovednosti... a dalších dovednostech můžeme... uvažovat jako o ,mentálních orgánech' [, které] se specifickým způsobem vyvíjí v souladu s genetickým programem... víceúčelové učící strategie přestanou existovat, budou pouze obecné principy ,růstu orgánů' , které vysvětlují tvar, strukturu a růst ledviny.*“¹ (Chomsky, 1980, s. 138-139) Podle Chomskyho tak pojem učení ve spojitosti s osvojováním jazyka ztrácí svůj obvyklý význam a jako takový je tedy při jeho objasňování také naprosto neužitečný.

Platnost Goldovy a Chomskyho argumentace ve prospěch vrozenosti jazyka však oslabuje hned několik problematických předpokladů, ze kterých tato argumentace vychází (Elman, 1996): 1) První poněkud problematický předpoklad se týká toho, že znalost gramatiky, kterou si člověk osvojuje, je „dokonalá“ a bez chyb. Tento předpoklad se odráží také v Chomskyho důrazu na studium jazykové kompetence – předpokládaného souboru všech gramatických pravidel, která mluvčí jazyka (aniž by si toho byl vědom)² nosí ve své mysli, která jsou však vzhledem k omezeným kognitivním zdrojům mluvčího (jako je např. omezená paměťová kapacita) jen nedokonale realizována v rámci konkrétních jazykových promluv (tj. v rámci jazykové performance). Když připustíme, že člověk nikdy nedisponuje dokonalou a bezchybnou znalostí gramatiky (pro což ostatně ani neexistují žádné přímé důkazy) a že jeho reálná znalost gramatiky se té ideální pouze blíží, tvrzení, že z nedokonalého vstupu nikdy nelze získat (ne)dokonalý výstup, tak ztrácí na své „fatálnosti“. 2) Druhý problematický bod spočívá v tom, že Chomsky se až příliš soustředí na vrozené aspekty jazyka a ignoruje možnost, že by mohly existovat četné zdroje nepřímě dostupných negativních příkladů, které by děti mohly využívat při osvojování gramatiky i v situaci, kdy se v jazykovém prostředí nachází jen nedokonalá a neúplná vstupní data. Takto si lze například velice dobře představit, že dítě ve své mysli (nevědomě) konfrontuje svá očekávání ohledně toho, co kdo řekne v jaké situaci, s tím, co nakonec ve skutečnosti slyší.³ 3) Z hlediska tématu probíraného v tomto oddílu se jako nejdůležitější omezení Goldovy a Chomskyho argumentace jeví být jejich předpoklad ohledně povahy učícího mechanismu, který má údajně „trpět“ nedostatkem schopností potřebných k osvojení jazyka. Tento učící mechanismus má v Chomskyho pojetí podobu poměrně těžkopádného a biologicky a také psychologicky nepříliš plausibilního

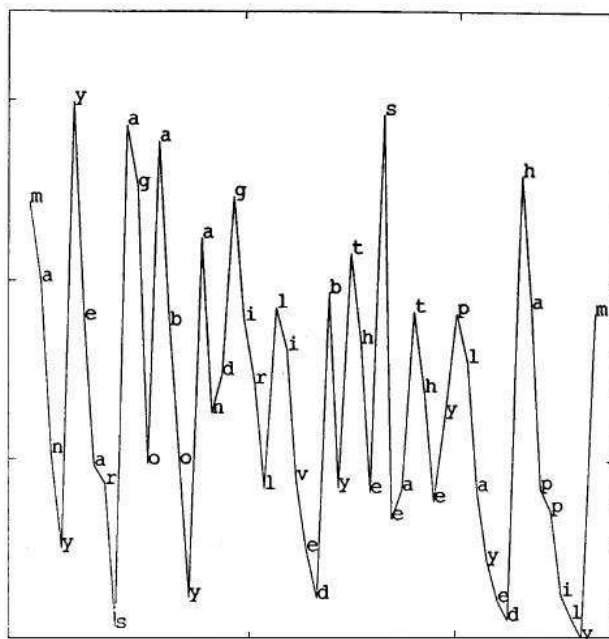
¹ „*We may usefully think of the language faculty, the number faculty, and others, as ,mental organs' [that] develop in specific ways, each in accordance with the genetic program... multipurpose learning strategies are no more likely exist than general principles of ,growth of organs' that account for the shape, structure and growth of the kidney.*“

² Tzv. „tichá znalost“, *tacit knowledge*.

³ Na této myšlence je také založena predikční úloha, kterou plní Elmanova jednoduchá rekurentní síť při odhalování gramatické struktury jazyka (více viz oddíl „3.2.3.5.4 Konekcionistický model distribuční analýzy“).

procesu generování a testování hypotéz. Tento učící mechanismus skutečně může mít potíže s osvojováním jazyka. Je to však stejné jako tvrdit, že člověk, který při autonehodě přišel o obě dolní končetiny, může mít potíže s tím, dostat se na vrcholek Mount Everestu: Stejně jako autonehoda může ve skutečném životě připravit člověka o prostředky nutné ke zdolání nejvyšší hory světa, tak Chomsky ve své teorii připravil člověka o prostředky potřebné k osvojení mateřského jazyka. Je jenom příznačné, že Chomskyho obraz jazyka nachází své (výpočetní) modely v tradičních symbolových systémech, které se jazyku neučí, ale je jim spíše dopředu naprogramován. Biologicky a psychologicky mnohem plausibilnějším modelem učení je umělá neuronová síť - specifický typ (paralelně distribuovaného) výpočetního zařízení, které využívá poměrně jednoduchých a univerzálních učících algoritmů. Četné práce v oblasti konekcionistického modelování jazykového vývoje ukázaly, že umělé neuronové sítě dokáží i z neúplných vstupních dat vyextrahovat statistické vzorce a pravidelnosti související s různými jazykovými jevy a generalizovat takto získané znalosti i na nové a dosud nepotkané případy. Jako příklad zde lze uvést schopnost jednoduchých rekurentních sítí na základě statistických zákonitostí v (sou)výskytu jednotlivých hlásek detekovat ve spojitém toku řeči hranice jednotlivých slov jako místa, kde je podmíněná pravděpodobnost výskytu další hlásky za předpokladu výskytu všech předcházejících hlásek velice nízká (Elman, 1990; viz také obrázek 161). Jiným příkladem je schopnost stejného typu sítě si osvojit znalost některých gramatických a sémantických kategorií čistě na základě statistických zákonitostí ve výskytu jednotlivých slov v různých jazykových kontextech (Elman, 1990, 1993). Na základě jazykových vstupů běžně dostupných malým dětem si podle stejného principu jednoduchá rekurentní síť také dokáže osvojit gramatickou strukturu jazyka, která je nutná ke zvládnutí tvorby otázek, o kterých Chomsky (1975) předpokládal, že se je dítě nikdy nemůže naučit čistě jenom na základě empirické evidence dostupné v jeho jazykovém prostředí (Lewis, Elman, 2002). Plausibilitu těchto konekcionistických modelů potvrzují také některé empirické studie statistického učení v souvislosti se schopností malých dětí segmentovat spojitý tok řeči na jednotlivá slova (Saffran, Aslin, Newport, 1996) a vyvozovat gramatická pravidla jednoduchého umělého jazyka (Gomez, Gerken, 1999) na základě statistických pravidelností ve výskytu jednotlivých hlásek, slabik nebo slov. Práce s umělými neuronovými sítěmi v rámci konekcionistického modelování jazykového vývoje takto dokládají, že i v neúplném jazykovém vstupu je obsaženo dostatek informací, které dostatečně „mocný“ učící mechanismus jako jsou umělé neuronové sítě dokáže využít k extrakci zákonitostí, které jsou důležité z hlediska osvojení jazykové kompetence. Tzn. že

při rozhodování, co se z jazyka lze naučit a co již nikoli, je potřeba být velice opatrný - a umělé neuronové sítě mohou být při tomto rozhodování velice užitečnou pomůckou.



Obrázek 161: Vzorec individuálních chyb jednoduché rekurentní sítě predikující svůj příští vstup, kterým je vždy jedna z hlásek, které tvoří spojité tok řeči: „many years ago a boy and girl lived by the sea they played happily...“. Hranice jednotlivých slov je v grafu vyznačena relativními chybovými maximy: Čím dále od počátku slova, tím je zjištěná chyba ve výstupní vrstvě menší, což lze interpretovat jako postupně stoupající jistotu neuronové sítě ohledně toho, jaká další hláska bude v řečovém projevu následovat; na počátku nového slova je zjištěná chyba ve výstupní vrstvě relativně nejvyšší, neboť po jednom slově může většinou následovat velký počet různých slov. (převzato z Elman, 1990)

To, že se toho umělé neuronové sítě dokáží na základě vlastních zkušeností tolik naučit, však ještě neznamená, že by mechanismus učení byl všemocný a že by na jeho základě bylo možné všechno vysvětlit. V této souvislosti se často objevují obavy nebo námitky, že konekcionismus do psychologie a dalších příbuzných disciplín (znovu)zavádí zdánlivě již dávno pohřbený koncept lidské mysli jako nepopsané desky, na kterou se zaznamenávají otisky všech zkušeností člověka se světem. Skutečnost je taková, že i jen nepatrná změna v některých parametrech neuronové sítě může rozhodnout o tom, zda si neuronová síť dokáže, nebo nedokáže požadovanou funkci osvojit. Nastavení parametrů přitom (v drtivé většině případů) není pod kontrolou samotné neuronové sítě, ale je jí „vnuceno“ zvnějšku experimentátorem, který je často nucen velice dlouho hledat tu správnou neuronovou síť s tou správnou architekturou, aktivační funkcí, mírou učení..., kterou by se mu podařilo na základě prezentovaných příkladů přimět chovat se požadovaným způsobem. Umělá neuronová síť tak

„přichází na svět“ již s určitými vrozenými dispozicemi v podobě architektury sítě, počtu uzlů ve skryté vrstvě nebo nastavení některých jejích klíčových parametrů jako je typ aktivační funkce, parametr míry učení apod. Tyto vrozené dispozice však mají poněkud odlišnou povahu než ty, se kterými se lze setkat v psycholingvistice silně ovlivněné Chomskyho koncepcí univerzální gramatiky: Zatímco v případě umělých neuronových sítí mají vrozené dispozice povahu spíše obecných, doménově nespecifických výpočetních omezení či principů, v případě Chomskyho univerzální gramatiky se jedná o doménově vysoce specifické poznatky týkající se některých konkrétních vlastností jazyka. V tabulce 11 jsou přehledně uvedeny hlavní zdroje (vrozených) omezení, které se mohou uplatňovat při zpracovávání informací v biologických nebo umělých neuronových sítích. Tato omezení se nacházejí podél škály, která rozlišuje jednotlivé zdroje omezení podle míry jejich doménové ne/specifity, resp. bezprostřednosti, s níž zpracovávání informací ovlivňují – zatímco doménově nejvíce specifické jsou reprezentace konkrétních poznatků v podobě specificky nastavených vah (synaptických) spojů, doménově nejméně specifická jsou časová omezení různých vývojových nebo tréninkových procesů odehrávajících se v mozku nebo v umělé neuronové síti. Z údajů v tabulce je zřejmé, že znalosti (například jazyka) nemusí být vrozené přímo, ale že může být vrozený pouze mechanismus, který se tyto znalosti dokáže snadno a velice rychle naučit. Manifestací takového vrozeného učícího mechanismu jsou zde již několikrát zmiňovaná kritická období ve vývoji řady živočišných druhů (včetně člověka), během kterých se mláďata daného druhu dokáží velice rychle naučit určitému specifickému druhu chování na základě jen velice krátké a omezené zkušenosti s odpovídajícím „podnětovým materiálem“. Vznik takového učícího mechanismu prostřednictvím evoluce přírodním výběrem ve své práci simulovali Nakisa a Plunkett (1998), kterým se s pomocí genetických algoritmů (kdy do populace umělých „genomů“ zakódovali architekturu a učící algoritmus sítě a poté tuto populaci vystavili iterativnímu procesu kumulativní selekce) podařilo „vyšlechtit“ konekcionistický model osvojování senzitivity vůči hláskové struktuře jazyka na bázi atraktorové sítě, která se dokázala během pouhých dvou minut expozice řečovým podnětům naučit kategorizovat tyto řečové podněty v souladu s hláskovou strukturou anglického jazyka. Výsledkem umělého výběru zde tak nebyla konkrétní znalost hláskové struktury anglického jazyka, ale neuronová síť s tou správnou architekturou a s tím správným učícím algoritmem, která se dokázala velice rychle a snadno naučit požadovanému chování. V tomto kontextu je také zajímavá Elmanova práce z roku 1993, ve které její autor prokázal, že i doménově nespecifické omezení v podobě inkrementální prezentace tréninkových dat může mít doménově vysoce specifické důsledky (více viz s. 320-321).

Zdroj omezení		Příklady v mozku	Příklady v UNS
<div style="writing-mode: vertical-rl; transform: rotate(180deg);"> Nejméně specifická/nepřímá ↓ Nejvíce specifická/přímá </div>	Reprezentace	synapse; specifické mikroobvody	váhy spojů
	Architektura		
	jednotka	cytoarchitektura (typy neuronů); práh excitace; druhy neurotransmiterů; učící pravidla (například LTP)	aktivační funkce; učící algoritmus; parametr momentum; parametr míry učení
	lokální	počet vrstev; hustota; rekurence; základní kortikální obvody	typ sítě (například rekurentní, dopředná); počet vrstev, počet uzlů v jednotlivých vrstvách
	globální	spoje mezi jednotlivými mozkovými oblastmi	expertní sítě; oddělené vstupní/výstupní kanály
	Načasování	počet dělení buněk během neurogeneze; prostoro-časové vlny synaptického růstu a prořezávání; časový vývoj senzorických systémů	inkrementální prezentace dat; dělení uzlů v rostoucí síti; adaptivní míra učení; vnitřní změny způsobené sycením jednotlivých uzlů

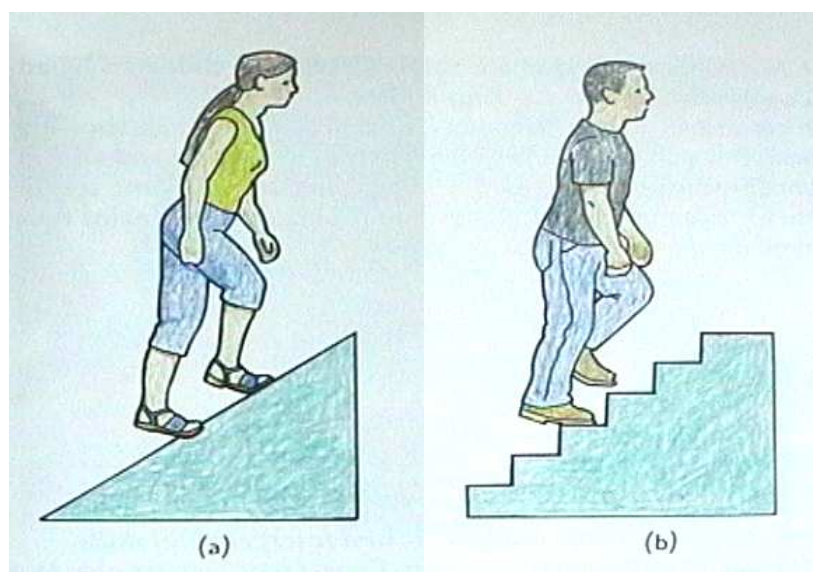
Tabulka 11: Přehled možných zdrojů vrozených omezení uplatňujících se při zpracovávání informací v biologických a umělých neuronových sítích. (převzato z Elman a kol., 1996)

3.2.4.2 Kontinuální vs. stupňovitý vývoj

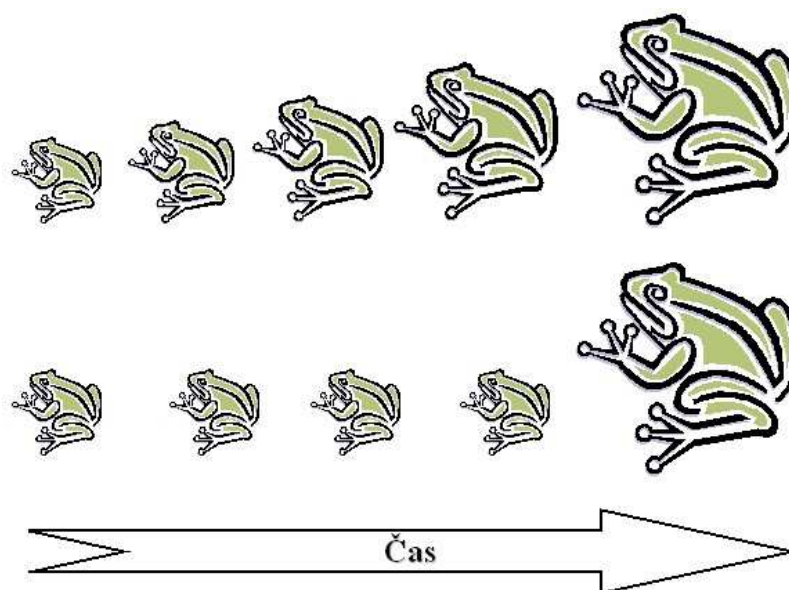
Kromě toho, že konekcionistické modely umožňují prakticky ověřit, co všechno se lze naučit na základě vlastní zkušenosti, resp. na základě vstupů z vnějšího prostředí, tento typ výpočetních modelů značně přispívá také k osvětlení možných mechanismů, které mohou stát za specifickým „tvarem“ vývojových změn, ke kterým dochází v průběhu vývoje různých kognitivních funkcí.

Jak známo, kognitivní vývoj má jen málokdy podobu jednoduchého lineárního růstu či monotónního zvyšování kompetence v dané kognitivní doméně. Ve vývoji kognitivních funkcí lze naopak velice často pozorovat etapy zrychleného nebo naopak zpomaleného vývoje, tzv. kritická období, která jsou optimální pro rozvoj určitých kognitivních funkcí, náhlé objevení se nových schopností, ale stejně tak jejich dočasnou ztrátu nebo zhoršení, tedy přesný opak toho, co by člověk očekával od (progresivního) vývoje směřující od méně schopného a kompetentního dítěte ke stále schopnějšímu a kompetentnějšímu dospělci.

Jeden z nejdůležitějších aspektů problematiky vývojových mechanismů skrývajících se za profilem vývojových změn má podobu sporu o to, zda má kognitivní vývoj kontinuální nebo naopak diskrétní (stupňovitou) povahu, tedy zda probíhá spíše v podobě drobných kroků a pozvolných změn nebo naopak v podobě náhlých přechodů mezi kvalitativně zcela odlišnými vývojovými stádii, které přicházejí po období relativního (vývojového) klidu a stagnace. Pod pojmy *punktualismus* a *gradualismus* se obdobná debata vede také v souvislosti s fylogenetickým vývojem organismů (viz obrázek 163; Dawkins, 2002).



Obrázek 162: Ve vývojové psychologii se vede spor o to, zda kognitivní vývoj má povahu kontinuálně probíhajících drobných změn (a), anebo zda v kognitivním vývoji existují období relativního klidu (kdy se schopnosti dítěte nijak dramaticky nemění), která jsou od sebe oddělena určitými „kritickými body“, kdy dochází k náhlému (skokovému) přechodu z jedné kvalitativně odlišné vývojové fáze do druhé (b).



Obrázek 163: Zatímco podle gradualistů se evoluce organismů děje prostřednictvím kumulace drobných změn v průběhu dlouhých časových období (horní část obrázku), podle punktualistů k evolučním změnám dochází náhle, ve skocích, de facto během jediné generace (dolní část obrázku).

Ve vztahu k této otázce má na současnou vývojovou psychologii stále velký vliv švýcarský psycholog *Jean Piaget* a jeho koncepce stádií kognitivního vývoje (Piaget, Inhelderová, 1970; Piaget, 1999). Jakkoli byly zpochybněny četné Piagetovy závěry ohledně věkových hranic jednotlivých stádií kognitivního vývoje, ohledně toho, co děti v kterém věku dokáží a co nikoli, a také ohledně konkrétních příčin selhání dětí v jednotlivých piagetovských úlohách (viz například Spelke, 1988, 1990; Baillargeon, 1993; Wynn, 1995; Thelen a kol., 2001), jeho základní myšlenka si i nadále zachovává značnou přitažlivost, kterou se zdají podporovat také některé z výše zmiňovaných charakteristik kognitivního vývoje. Základní Piagetovou myšlenkou je to, že člověk během svého kognitivního vývoje prochází řadou kvalitativně odlišných stádií, během kterých ve svém uvažování používá (kvalitativně) zcela odlišné mentální procesy či operace. Podle Piageta například dítě ve svém nejranějším stádiu kognitivního vývoje disponuje pouze ne-symbolickými operacemi, které pracují jen s těmi informacemi, které jsou dítěti bezprostředně dostupné prostřednictvím jeho smyslů. Důsledkem je pak to, že dítě v této fázi kognitivního vývoje není schopné řešit úlohy, které vyžadují schopnost mentálně manipulovat s informacemi, které nejsou v prostředí aktuálně přítomné nebo nejsou bezprostředně smyslově dostupné. Jakmile se však u dítěte v následujícím vývojovém stádiu objeví kvalitativně nový kognitivní mechanismus v podobě symbolických reprezentací, dítě najednou začne být schopné úspěšně plnit i ty úlohy, které v předcházejícím vývojovém stádiu schopné řešit nebylo.

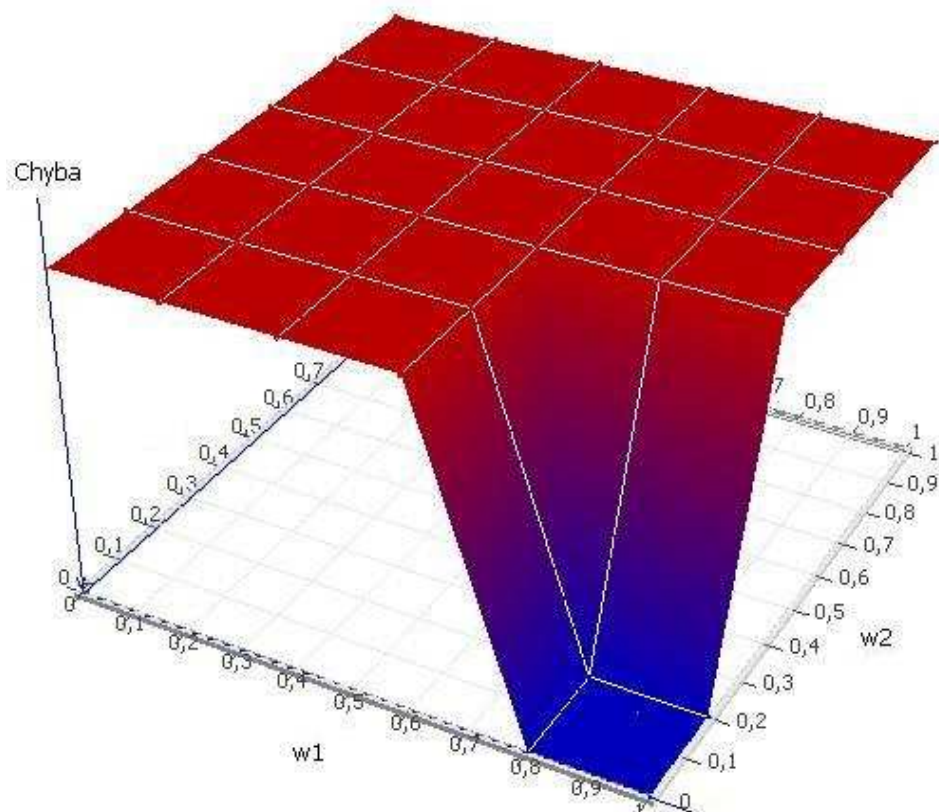
Piaget k mapování vývojových stádií vymyslel řadu experimentálních situací, v rámci kterých děti řešily různé kognitivní úlohy, a podle toho, jakých chyb se děti v průběhu jejich řešení dopouštěly, Piaget usuzoval na skryté mentální mechanismy stojící za pozorovatelným chováním. Jedna taková úloha, s jejíž pomocí Piaget zkoumal vývoj schopnosti decentrace a reverzibility myšlení, spočívala v posuzování, na kterou stranu (a jestli vůbec) se přikloní dvouramenná váha v závislosti na různě velkých závažích umístěných na dvou ramenech v různé vzdálenosti od otočného bodu (tzv. *balance beam problem*). Nověji výkon dětí v této úloze zkoumal Siegler (1981), který zjistil, že děti při řešení této úlohy prochází čtyřmi různými vývojovými stádii, během kterých se řídí – nebo se to vnějšímu pozorovateli alespoň tak zdá, že se řídí – čtyřmi různými pravidly, která v různé míře (a za různých podmínek) zohledňují dva základní aspekty problémové situace, které jsou relevantní z hlediska jejího řešení.¹ Taková zjištění mohou sugerovat představu, že tyto čtyři různé vývojové fáze v chování dítěte odrážejí kvalitativní změny ve způsobu, jakým dítě zpracovává, resp. reprezentuje informace o vnějším prostředí.

Avšak již v jednom z prvních pokusů o konekcionistické modelování kognitivního vývoje McClelland (1989) vytvořil poměrně jednoduchou dopřednou neuronovou síť, která úspěšně simulovala vývoj výkonu dítěte v úloze s dvouramennou vahou - tzn. že chování této neuronové sítě v průběhu jejího tréninku vykazovalo stejný stupňovitý vývoj, jaký je možné pozorovat u dětí. A vzhledem k tomu, že je velice dobře známo, jakým způsobem se neuronová síť učí a vyvíjí - totiž prostřednictvím postupných a drobných úprav vážených spojení a bez nějaké zásadní změny samotné architektury sítě nebo učicího algoritmu -, lze z úspěchu McClellandovy sítě při simulaci stupňovitého vývoje celkem bezpečně usuzovat na to, že i malé, pozvolné a kontinuální změny (ke kterým dochází při úpravě vah spojení učící se neuronové sítě) mohou mít za následek náhlé objevení se kvalitativně zcela nového způsobu chování.² Tzn. že k vysvětlení takového kvalitativně zcela nového způsobu chování není potřeba vždy automaticky předpokládat také existenci nebo vznik kvalitativně zcela nového mentálního mechanismu. To, jak je něco takového možné, lze jednoduše ilustrovat s pomocí

¹ V první fázi děti věnují pozornost pouze velikosti závaží; jeho vzdálenost od osy otáčení zcela ignorují a neberou ji nijak v úvahu. Ve druhém stádiu děti zohledňují také vzdálenost od osy otáčení, ale pouze v tom případě, že závaží na obou ramenech váhy jsou stejně velká. Ve třetí fázi děti dokáží správně zohlednit oba dva aspekty problémové situace, ale pouze tehdy, když je druhý aspekt situace na obou ramenech váhy totožný. Konečně ve čtvrtém stádiu již děti implicitně pracují s konceptem točivého momentu a dokáží náležitě zohlednit oba dva aspekty problémové situace, a to i tehdy, když jsou tyto dva aspekty v rozporu (tedy když na jednom rameni je větší závaží a na druhém se závaží nachází ve větší vzdálenosti od osy otáčení).

² Což samozřejmě ještě nic neříká o tom, že stejně je tomu i u člověka, nicméně vzhledem k tomu, že umělé neuronové sítě na určité rovině abstrakce fungují na základě stejných principů jako biologické neuronové sítě, je potřeba být přinejmenším značně opatrný při postulování existence kvalitativně zcela odlišných kognitivních mechanismů čistě na základě zjištěných behaviorálních dat.

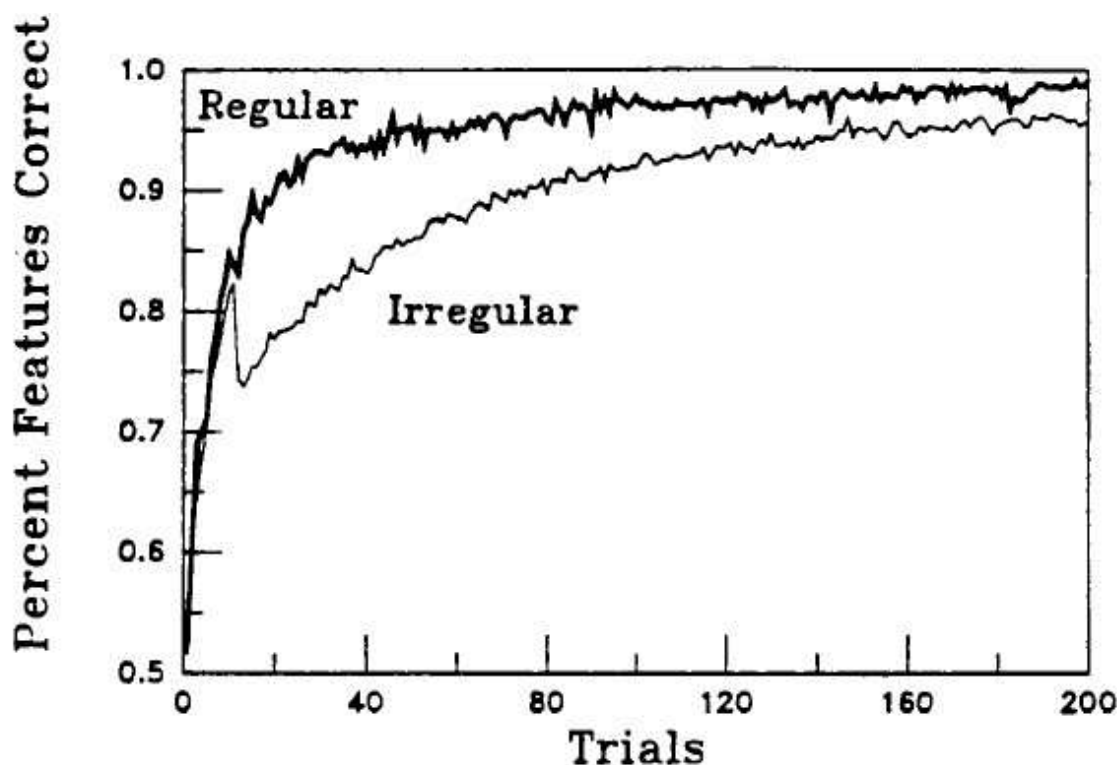
obrázku 164, na kterém je zachycena část chybové krajiny jednoduché neuronové sítě se dvěma váženými spoji: Na obrázku je dobře vidět, že díky nelinearitě obsažené v aktivační funkci (viz s. 216-217) i nepatrná změna v hodnotě jednoho váženého spoje (například zvýšení váhy spoje w_1 z 0,6 na 0,8) může vést k náhlému a prudkému snížení celkové chyby ve výstupní vrstvě a ke stejně náhlému a prudkému objevení se odpovídající kognitivní funkce, kterou tato jednoduchá umělá neuronová síť implementuje.



Obrázek 164: Chybová krajina - stavový prostor umělé neuronové sítě definovaný jejími dvěma váženými spoji a mírou celkové chyby ve výstupní vrstvě. I minimální změna v hodnotě jednoho váženého spoje zde může díky nelinearitě obsažené v aktivační funkci vést k výraznému snížení celkové chyby ve výstupní vrstvě a v důsledku toho také k náhlému objevení se kognitivní funkce, kterou neuronová síť implementuje.

Rumelhartovi s McClellandem (1986d) se podobně podařilo s pomocí dopředné (dvouvrstevné) neuronové sítě vytvořit konekcionistický model úspěšně simulující vývojová stadia, kterými prochází děti z anglofonního prostředí při osvojování schopnosti vytvářet minulý čas od pravidelných a nepravidelných sloves. Tuto schopnost si děti osvojují ve třech základních fázích: 1) V první fázi děti správně vytvářejí minulý čas jak od pravidelných sloves (přidáním koncovky *-ed* k základnímu tvaru slovesa v přítomném čase) tak i od sloves nepravidelných. 2) Ve druhé fázi děti vytvářejí minulý čas od nepravidelných sloves stejným

způsobem jako od pravidelných sloves (přidáním koncovky *-ed*). 3) A ve třetí fázi děti opět tvoří správně tvary jak od pravidelných tak i nepravidelných sloves. Vývojová křivka tak má charakteristický tvar písmene U, kdy po počátečním dobrém výkonu dochází k jeho dočasnému zhoršení, po kterém je opětovně dosaženo původní úrovně výkonu. Podle tradičního vysvětlení jsou příčinou této specifické vývojové křivky kvalitativně odlišné kognitivní (reprezentační) mechanismy, které se uplatňují v jednotlivých fázích jazykového vývoje. Zatímco v první fázi je schopnost dítěte vytvářet minulý čas od pravidelných a nepravidelných sloves založena čistě na mechanickém zapamatování si tvarů sloves v minulém čase, ve druhé fázi si dítě na základě své zkušenosti s anglickým jazykem (ve kterém převažují pravidelná slovesa) osvojuje abstraktní pravidlo pro vytváření minulé času, (které říká *cosi ve smyslu Vem základní tvar slovesa v přítomném čase a přidej k němu koncovku -ed*) a aplikuje ho na všechna slovesa, a to bez ohledu na to, zda se jedná o slovesa pravidelná nebo nepravidelná (tzv. *over-regularization error*, tj. *chyba nadměrného užívání gramatických pravidel*). Ve třetí fázi si pak dítě uvědomí, že z tohoto pravidla existují četné výjimky, a začne při vytváření minulé času rozlišovat mezi pravidelnými a nepravidelnými slovesy. Na konci celého vývoje tak dítě disponuje dvěma různými mentálními mechanismy, které mají na starosti vytváření minulé času od dvou různých skupin sloves (tzv. *dual-route model*). Rumelhartův a McClellandův výpočetní model však prokázal, že stejné vývojové křivky (ve tvaru písmene U) lze dosáhnout také u umělé neuronové sítě, jejíž architektura a učicí algoritmus se v průběhu učení nijak nemění. Fáze, kterými neuronová síť během svého vývoje procházela, byly způsobeny čistě interakcí neměnného učicího mechanismu s (tréninkovým) prostředím majícím určitou specifickou statistickou strukturu. Právě ta se stala terčem kritiky Pinkera a Prince (1988), kteří upozorňovali na to, že zastoupení pravidelných a nepravidelných sloves v tréninkovém prostředí neuronové sítě (resp. jejich „dávkování“ v průběhu jejího tréninku) neodpovídalo podmínkám, za kterých se tvorbě minulé času sloves učí malé děti: Podle Pinkera a Prince se v prvních třech letech života ve slovníku dětí (z anglofonního prostředí) nachází zhruba stejný počet pravidelných i nepravidelných sloves; v tréninkovém prostředí Rumelhartovy a McClellandovy neuronové sítě však na samotném počátku tréninku (v prvních deseti epochách) značně převládala nepravidelná slovesa, s tím, že později se poměr pravidelných a nepravidelných sloves obrátil ve prospěch početnějších pravidelných sloves. To vedlo k podezření, že stádia ve vývoji jazykové kompetence neuronové sítě byla způsobena čistě umělým (a vývojovými daty nijak



Obrázek 165: Výkon Rumelhartovy a McClellandovy neuronové sítě při tvorbě minulého času od ne/pravidelných sloves v závislosti na uběhnutých epochách tréninkového procesu. Z grafu je dobře patrné, jak okolo desáté tréninkové epochy došlo k prudkému snížení schopnosti neuronové sítě správně tvořit minulý čas od nepravidelných sloves. V tomto bodě neuronová síť začala vykazovat chování popisované v psycholingvistické literatuře jako chyba nadměrného užívání gramatických pravidel (*over-regularization error*). (převzato z Rumelhart, McClelland, 1986d)

nepodloženým) zásahem do struktury tréninkového prostředí.¹ Plunkett a Marchman (1991, 1993, 1996) tento nedostatek v pozdější simulaci odstranili a ukázali, že umělá neuronová síť (tentokrát i se skrytou vrstvou) vykazuje typickou vývojovou křivku a typický vzorec chyb i bez podobných umělých zásahů a že pozorovaná stádia v chování umělé neuronové sítě jsou dána čistě interferencí mezi dvěma odlišnými typy sloves při úpravách váhové matice v průběhu trénování sítě.

¹ Na počátku tréninku bylo v tréninkovém prostředí pouze deset sloves, z nichž osm bylo nepravidelných a dvě pravidelná. Vzhledem k obrovskému počtu vážených spojů nebyl pro neuronovou síť žádný problém naučit se minulý čas od všech deseti sloves mechanicky nazpaměť. V další fázi tréninku začaly v tréninkovém prostředí převládat pravidelná slovesa, takže především tento typ sloves měl vliv na úpravu vah spojů v průběhu trénování sítě. Neuronová síť s takovou váhovou maticí měla tendenci zpracovávat i nepravidelná slovesa jako slovesa pravidelná, což mělo automaticky za následek efekt tradičně připisovaný chybě nadměrného užívání gramatických pravidel (*over-regularization error*); viz také obrázek 165.

3.2.4.3 Způsob reprezentace poznatků

Rumelhartova a McClellandova simulace také dokazuje, že chování, které z vnějšku působí jako důsledek dodržování nějakého abstraktního a explicitního (přitom ale ne nutně člověkem uvědomovaného) pravidla, nemusí být na znalosti takového pravidla vůbec založeno. To samozřejmě neznámá, že by se neuronové sítě neřídily žádnými pravidly - ony se jimi řídí (koneckonců umělé neuronové sítě aproximují různé funkce a funkce nejsou ničím jiným než druhem pravidla), ale jsou to poněkud jiná pravidla než jsou ta, se kterými se lze setkat v tradičních symbolových systémech nebo v lingvistických analýzách: Na rozdíl od nich se jedná o pravidla, která jsou definována na objektech, které mají hodně daleko k sémantice běžného (lingvistického) diskurzu operujícího s takovými pojmy jako jsou hlásky, morfémy, slovní druhy, větné členy, frázové struktury apod. (více viz oddíl „1.2.3.4 Výpočetní architektura mysli/mozku,“). Konekcionistické modely tímto přispívají ke značnému obohacení našich představ o tom, jak mohou být různé poznatky reprezentovány a co to vlastně znamená, když se řekne, že někdo něco ví a že disponuje určitými znalostmi.

3.2.4.4 Kritická období

Za doklad toho, že člověk ve svém kognitivním vývoji prochází řadou stádií, během kterých se jeho kognitivní systém řídí kvalitativně zcela odlišnými principy, se považuje také existence tzv. „kritických období“, kdy se zdá, že dítě je po jistou omezenou dobu vybaveno schopnostmi, které mu umožňují si relativně snadno a rychle osvojovat různé kognitivní funkce. Po uplynutí určité doby však dítě tyto schopnosti nenávratně ztrácí a vyskytne-li se v nesprávnou dobu na nesprávném místě, kde se mu z vnějšího prostředí nedostane odpovídající stimulace, brány k jazykům a dalším kognitivním funkcím se mu tím nadobro uzavřou, budeme-li volně parafrázovat název jednoho z děl Jana Ámose Komenského. Práce s umělými neuronovými sítěmi využívajícími logistickou aktivační funkci (a učící algoritmus zpětného šíření nebo pravidlo nejmenších čtverců) však ukazují, že „kritická období“ se ve vývoji umělých neuronových sítí objevují i beze změny samotného učícího mechanismu, tj. beze změny architektury sítě a jejího učícího algoritmu. Ztráta plasticity takové neuronové sítě (tj. její schopnosti se dál učit) je čistě důsledkem specifického tvaru logistické aktivační funkce, jejíž sklon má maximální hodnotu tam, kde se celkový vstup ($netinput_i$) do uzlu i pohybuje okolo 0, a čím více se od této hodnoty vzdalujeme, tím je sklon aktivační funkce menší (viz obrázek 123 na straně 217). To vede k tomu, že maximální míra učení (tj. největší

změny vah spojů) probíhá u uzlů, jejichž celkový vstup se pohybuje okolo 0, tzn. u těch uzlů, kterým ještě nebyla přiřčena určitá konkrétní role v procesu transformace vstupů na výstupy; naproti tomu uzly s velkým celkovým vstupem, které jsou již plně zapojeny do „transformačního procesu“, mění váhy svých spojů už jen minimálně. Tzn. že učení neuronové sítě probíhá pouze v určitém omezeném časovém období, kdy dochází k postupnému „usazování rolí“ jednotlivých uzlů sítě a po jehož uplynutí je už jen velice obtížné toto nastavení jakkoli měnit. Neuronová síť takto ztrácí svoji plasticitu v důsledku samotného procesu učení, aniž by přitom docházelo k jakékoli kvalitativní změně principů, které stojí v pozadí jejího fungování.

3.2.4.5 Operacionalizace vývojově-psychologických pojmů

Vedle toho, že konekcionistické výpočetní modely nabízejí nový úhel pohledu na tradiční a zavedené pojmy vývojové psychologie, dovolují také dát trochu konkrétnější a více hmatatelnou podobu jinak jen velice vágně definovaným pojmům, se kterými se v rámci různých vývojově-psychologických teorií často pracuje. Jako příklad zde můžeme uvést Piagetovy pojmy *asimilace* a *akomodace*, dva základní mechanismy kognitivního vývoje: Podle Piageta si dítě na základě svých zkušeností ve své mysli vytváří různá *kognitivní schémata* či *mentální rámce*, ve kterých jsou uloženy veškeré znalosti dítěte o tom, jak to „chodí“ v jeho fyzikálním a sociálním prostředí. Funkcí takto vytvořených schémat je *adaptace na prostředí*, kterou Piaget chápe jako záležitost rovnovážného stavu (*ekvilibria*) mezi prostředím a kognitivními schématy, jejichž prostřednictvím dítě zpracovává podněty, které k němu z tohoto prostředí přicházejí. K ustavování této rovnováhy dochází podle Piageta na základě dvou základních kognitivních mechanismů, kterými jsou *asimilace* a *akomodace*: Nové jevy se člověk vždy snaží začleňovat do již existujících a dostupných schémat, která si vytvořil na základě svých minulých zkušeností (proces *asimilace*). Jestliže jsou však tato stávající schémata neadekvátní novým podnětům, vede to ke vzniku jakési *kognitivní nerovnováhy*, která dítě nutí odstranit existující nesoulad mezi poznáním a realitou. Opětovné nastolení rovnováhy se děje prostřednictvím mechanismu *akomodace*, kdy dítě vytváří nová schémata nebo stávající schémata modifikuje tak, aby do nich nové jevy mohlo začlenit. Neustálým střídáním stavů nerovnováhy (vyvolaných novými zkušenostmi) a stavů opětovně nabyté rovnováhy (dosažených vytvořením nových schémat vyhovujících všem podstatným informacím z prostředí) dochází k postupnému propracovávání kognitivních schémat umožňujících efektivnější a adaptivnější organizování zkušeností s prostředím.

V rámci konekcionistického modelování lze proces asimilace chápat jako stav, kdy neuronová síť ke zpracovávání podnětů přicházejících k ní z vnějšího prostředí přes její vstupní vrstvu využívá stávající váhovou matici, která byla nastavena na základě všech jejích dosavadních zkušeností. Proces akomodace by pak odpovídal stavu, kdy se nastavení vah jednotlivých spojů mezi uzly sítě mění pod vlivem nových zkušeností. S takto operacionalizovanou definicí mechanismů asimilace a akomodace je již mnohem snazší podrobit případnému testování některé z hypotéz, které vyplývají z Piagetova modelu kognitivního vývoje. Počítačová implementace takového modelu by rovněž umožňovala generovat nové hypotézy, neboť bez pomoci počítače lze často jen velice obtížně předvídat možné důsledky komplexních interakcí mezi jednotlivými prvky modelu.

S pomocí umělých neuronových sítí lze podobně precizovat také další pojmy jako je například pojem *vrozených dispozic* (viz oddíl „3.2.4.1 Nativismus vs. empirismus“) nebo pojem *emergence*. Ten může v rámci konekcionistického modelování označovat předem nenaprogramované chování, které vyvstává z interakce mezi obecnými principy zpracovávání informací v umělé neuronové síti a specifickým způsobem prezentace vstupních dat z tréninkového prostředí s určitou specifickou statistickou strukturou. Jako konkrétní příklad emergentního chování umělé neuronové sítě zde lze uvést již jednou zmiňovaná kritická období ve vývoji umělé neuronové sítě či Elmanem (1993) objevený vliv inkrementálního způsobu učení na schopnost sítě si osvojit požadovanou kognitivní funkci.

SPECIÁLNÍ ČÁST

4 PDP model využití fonotaktických a prozodických vodítek k segmentaci spojitého toku řeči na jednotlivá slova v ČJ

„V rámci polárního... vnímání světa lidskou psychikou nejenže vydělujeme jednotlivé věci z v zásadě kompaktního a kontinuálního bytí, ale rozpadají se nám, ač i samy o sobě tvoří jeden celek, na mozaiku smyslových aspektů (vzhled, zvukové projevy, konzistence atd.) či jejich jemnější dělení (tvar, barva, lesk, témbra atd.) jako důsledek tohoto „zmocňování se“ světa pomocí naší psychy, které pochopitelně představuje na světě určité násilí - „krávolouka“ se rozpadá na krávu a louku zčásti až v důsledku našeho interpretačního úsilí, ale alespoň v určité míře „po letokruzích“ světa toto snažen jít musí [takže jen] těžko se najde jazyk, mající slovo pro souhrnné označení dobytka i s pastvinou a pro jednotlivou krávu ne. [...] Připomíná to známou čínskou metaforu o taoistickém řezníkovi, který dobytče rozporcovává podle přirozených anatomických struktur, takřka bez vynaložení energie na rozdíl od jeho nevědomého kolegy, který zvíře řeže hrubou silou hlava nehlava.“

STANISLAV KOMÁREK, *Příroda a kultura. Svět jevů a svět interpretací.*

4.1 Výzkumný problém

Jedním z prvních problémů, na které dítě narazí ve své snaze osvojit si svou mateřštinu, je lokalizace hranic jednotlivých slov, která se z pohledu (nebo lépe řečeno „z poslechu“) netrénovaného a neškoleného ucha malého dítěte zdají tvořit jeden nepřetržitý proud řečových zvuků, ve kterém nejsou na první „poslech“ patrné (resp. slyšitelné) žádné výrazné charakteristiky, na základě kterých by bylo možné usuzovat na to, kde jedno slovo končí a druhé začíná. Vypůjčíme-li si od Williama Jamese jeho slavnou „repliku“ ohledně struktury vědomého prožívání malého dítěte, potom ranou jazykovou zkušenost můžeme charakterizovat jako „bzučící zmatek“,¹ který (z hlediska lexikálního, nikoli však prozodického²) postrádá jakoukoli pevnější strukturu. Přestože je dospělý člověk za normálního stavu vědomí již nadobro a nenávratně „odstříhnut“ od svých vzpomínek na nejranější období svého života, i v dospělosti si lze velice snadno „uměle“ navodit zkušenost „primordiálního jazykového chaosu“, jehož fenomenologie se do značné míry podobá běžné zkušenosti s jakýmkoli cizím a neznámým jazykem, který netrénovanému mozku zní podobně

¹ „...buzzing confusion.“

² Viz oddíl „3.2.3.1 Senzitivita k prozodické struktuře jazyka“.

jako mateřština novorozenci, tj. jako souvislý a nijak nestrukturovaný tok řečových zvuků, ve kterém lze rozlišit pouze určitou specifickou melodii, nikoli však jednotlivá slova.¹

V oddíle věnovaném segmentaci řeči² jsem navrhl dívat se na snahu dítěte identifikovat v „bzučícím zmatku“ řečových podnětů jednotlivá slova jako na slézání strmé a zdánlivě dokonale hladké stěny, ve které se horolezec pokouší najít nějakou škvíru či trhlínu, kterou by mohl použít jako oporu při jejím slézání. V této analogii se mysl dítěte („jazykolezce“) snaží najít ve zdánlivě zcela náhodné posloupnosti řečových zvuků (hladké stěně) jisté zákonitosti (škvíry a trhliny), které by jí mohly pomoci lokalizovat v souvislém toku řeči hranice jednotlivých slov. Klíč k řešení problému segmentace řeči spočívá tedy v tom, že posloupnost řečových podnětů není zcela náhodná, ale že se řídí určitými zákonitostmi, které by lidská mysl mohla s pomocí vhodných analytických nástrojů odhalit a následně také využít při snaze dát strukturu a řád tomu, co okolo sebe slyší. Tyto zákonitosti se týkají především statistických zákonitostí v (sou)výskytu jednotlivých hlásek a určitých prozodických charakteristik jazyka (jako je například stálý přízvuk na první, poslední nebo předposlední slabice). Různé empirické výzkumy statistického učení u malých dětí (například Saffran, Aslin, Newport, 1996; Mattyše, Jusczyk a kol., 1999) prokázaly, že děti jsou skutečně vůči tomuto druhu statistických zákonitostí v jazykovém vstupu citliví a že znalost těchto zákonitostí dokáží využívat při segmentaci spojitého toku řeči na jednotlivá slova. Mysl malého dítěte tak musí být vybavena nějakým druhem statistického procesoru (*statistical learning device*), který dítěti umožňuje využívat toho, že (sou)výskyt jednotlivých hlásek a některých prozodických charakteristik jazyka není nahodilý, ale že se řídí určitými pravidly a zákonitostmi, která vytvářejí hláskovou a prozodickou strukturu daného jazyka.

Biologicky a psychologicky plausibilním modelem takového statistického procesoru je umělá neuronová síť. Ve své práci z roku 1990 Jeffrey Elman prokázal, že určitý specifický typ umělé neuronové sítě (tzv. jednoduchá rekurentní síť)³ se dokáže na základě expozice souvislému toku řeči úspěšně naučit identifikovat hranice jednotlivých slov jako ta místa ve spojitém proudu hlásek, kde je podmíněná pravděpodobnost výskytu další hlásky za předpokladu výskytu všech předchozích hlásek velice nízká, což lze formulovat i tak, že se jedná o takové místo v proudu řeči, kde (z hlediska umělé neuronové sítě) dochází ke ztrátě jakákoli souvislosti mezi danou hláskou a všemi předcházejícími hláskami. Podle Elmana tak

¹ Oproti novorozenci má však dospělý mluvčí v této situaci jednu podstatnou výhodu - ví totiž, že něco takového jako slova vůbec existuje a že jazykový vstup lze na tyto základní stavební jednotky rozložit (a že by tedy také bylo záhodno začít tato slova hledat a nedát si pokoj, dokud neučiní zadost své potřebě slyšet jazyk jako proud jasně od sebe odlišitelných slov).

² Viz oddíl „3.2.3.3 Segmentace spojitého toku řeči“.

³ Viz oddíl „2.2.2.3.2 Jednoduché rekurentní síť“.

umělá neuronová síť dokáže identifikovat hranice jednotlivých slov díky tomu, že je schopná rozeznat, kdy jí je znalost všech předcházejících hlásek tzv. „k ničemu“, neboť jí již nijak nepomáhá při predikci další hlásky. V tomto Elmanem navrženém konceptuálním rámci má svůj základ také PDP simulace popsaná v této kapitole, jejímž cílem je prozkoumat interakci fonotaktických a prozodických vodítek při segmentaci řeči a dát tak odpověď na otázku, **jak (a zda vůbec) se předpokládaná delimitativní funkce a „přidaná hodnota“ prozodického vodítka v podobě stálého přízvuku na první slabice projeví v činnosti umělé neuronové sítě při její snaze anticipovat svůj příští vstup, jejímž vedlejším produktem je informace o hranicích jednotlivých slov ve spojitém toku řeči.** Jinými slovy, cílem zde popsaného PDP modelu je zjistit, **zda Elmanem navržený konekcionistický model je schopen implementovat a tak i vysvětlit delimitativní funkci prozodického vodítka majícího podobu stálého přízvuku na první slabice.** V následujících oddílech jsou nejdříve popsány některé základní charakteristiky použitého PDP modelu jako je konkrétní podoba řešené úlohy, obsah tréninkového prostředí, způsob reprezentace vstupních a výstupních dat, architektura neuronové sítě či její tréninkové parametry; po tomto stručném představení PDP modelu (realizovaného s pomocí simulačního softwaru *Tlearn*)¹ následuje prezentace samotných výsledků a jejich diskuse z hlediska výše položené otázky.

4.2 Definování úlohy pro neuronovou síť

Tréninkové prostředí umělé neuronové sítě (dále UNS) sestávalo z deseti různých slov - šesti substantiv (kolega, student, svačina, hlava, slunce a prase), dvou adjektiv (vysoký a starý) a dvou sloves (dostat a potkat). Od každého substantiva se v tréninkovém korpusu nacházelo všech jeho sedm pádů, a to v množném i jednotném čísle. Adjektiva ve středním a ženském rodě (jednotného čísla) byla zastoupena pouze v prvním pádě; naproti tomu adjektiva v mužském rodě (jednotného čísla) byla zastoupena ve všech svých sedmi pádech. Obě slovesa byla v minulém čase, ve všech třech osobách jednotného čísla a v první osobě množného čísla. V případě substantiv, sloves a adjektiv v mužském rodě se každý jejich tvar v tréninkovém korpusu nacházel celkem třikrát; adjektiva v ženském a středním rodě pouze jednou. V tréninkovém prostředí se tak celkem nacházelo 299 jednotlivých slov.

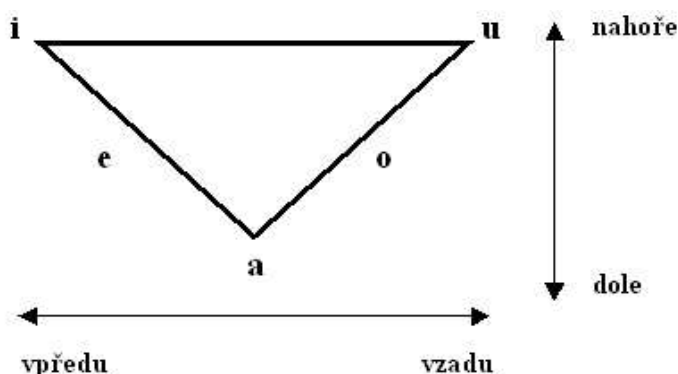
¹ Na přiloženém CD je v adresáři *Tlearn* umístěn instalační soubor tohoto simulačního softwaru spolu s manuálem k jeho ovládání. Případný zájemce tak má možnost si sám ověřit zde předkládané výsledky nebo experimentovat se cvičnými či svými vlastními simulacemi. Software (včetně manuálu) je volně k dispozici také na webových stránkách Jeffreyho Elmana, které jsou věnovány speciálně tomuto simulačnímu softwaru (<http://crl.ucsd.edu/innate/tlearn.html>).

S pomocí generátoru pseudonáhodných čísel bylo vytvořeno pořadí, v němž byla později jednotlivá slova prezentována UNS na její vstupní vrstvě. Ještě před samotnou prezentací však bylo každé slovo rozloženo na jednotlivé hlásky. V tréninkovém korpusu UNS se tak nenacházela slova, ale pouze spojitý tok 2 122 hlásek¹, ve kterém nebyly žádným způsobem vyznačeny hranice mezi jednotlivými slovy. Například v samotném úvodu tréninkového korpusu se nacházela následující ničím nepřerušovaná sekvence hlásek:

hlavastudentaprasevisokékolegoustarésluncíkolegistaréstarímvisokíkolegovi...

V každý okamžik byla UNS z této posloupnosti hlásek v pevném a neměnném pořadí prezentována právě jedna hláska. Úlohou UNS přitom bylo anticipovat svůj příští vstup, tj. hlásku, která následovala po té, která byla UNS aktuálně prezentována na její vstupní vrstvě. UNS úspěšně splnila svou úlohu tehdy, když dokázala v reakci na vstupní aktivační vzorec reprezentující určitou konkrétní hlásku vyprodukovat ve své výstupní vrstvě aktivační vzorec odpovídající následující hlásce v jazykovém vstupu. Použijeme-li jako příklad začátek výše uvedené části tréninkového korpusu, tak při percipování hlásky *h* by UNS měla být schopna ve své výstupní vrstvě anticipovat hlásku *l*.

Ve vstupní i výstupní vrstvě byly jednotlivé hlásky reprezentovány 26-ti místným vektorem jedniček a nul (viz obrázek 166). Toto reprezentační schéma vychází z klasifikace samohlásek a souhlásek, které ve své práci uvádí lingvista Jiří Černý (1998, s. 80-85): Zatímco klasifikace českých samohlásek je založena na jejich délce (dlouhé/krátké samohlásky) a na postavení jazyka při jejich artikulaci (přední/střední/zadní a vysoké/středové/nízké samohlásky; viz také obrázek 167), ke klasifikaci českých souhlásek se používají tři základní kritéria – způsob artikulace (závěrové/nosové/polozávěrové/třené/bo-



Obrázek 166: Hellwagův trojúhelník, který schematicky zachycuje postavení jazyka při artikulaci českých samohlásek. (s úpravami převzato z Černý, 1998)

¹ Soubor s posloupností těchto hlásek se nachází na přiloženém CD v adresáři *Hlásky_Exp* pod názvem *Exp1a*.

kové/kmitavé/hlasivkové), místo artikulace (obouretné/retozubné/dásňové přední/dásňové zadní/tvrddopatrové/měkkopatrové/hrtanové) a znělost (znělé/neznělé souhlásky).

prízvuk	samohláska	souhláska	znělost	přední	střední	zadní	vysoké	středové	nízké	dlouhé	krátké	závěrové	nosové	polozávěrové	třené	bokové	kmitavé	hlavivkové	oboustranné	retrozubné	dásňové přední	dásňové zadní	tvrdopatrové	měkkopatrové	hrtanové
a	0	1	0	1	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
ä	0	1	0	1	1	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
i	0	1	0	1	1	1	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
o	0	1	0	1	1	1	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
u	0	1	0	1	1	1	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
ä	0	1	0	1	1	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
é	0	1	0	1	1	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
i	0	1	0	1	1	1	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
ó	0	1	0	1	1	1	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
ú	0	1	0	1	1	1	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
p	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0
b	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0
t	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0
d	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0
ť	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0
k	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0
g	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0
m	0	0	1	1	1	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0
n	0	0	1	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0
ň	0	0	1	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0
č	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0
š	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0
ž	0	0	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0
j	0	0	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0

Obrázek 167: Reprezentační schéma použité ke kódování 35 hlásek českého jazyka.¹ Ve žlutých buňkách jsou zachyceny ty charakteristiky, které umožňují rozlišovat deset českých samohlásek. V šedivých buňkách jsou pak ty charakteristiky, které se vztahují k systému 25 českých souhlásek (s výjimkou znělosti, která se týká rovněž samohlásek, avšak nijak nepřispívá k jejich rozlišení, neboť samohlásky jsou všechny znělé). V bílých buňkách se pak nacházejí ty charakteristiky, které vstupní podněty klasifikují z hlediska ne/přítomnosti přízvuku a příslušnosti do množiny souhlásek/samohlásek.

¹ Z čistě praktických důvodů (souvisejících s převodem jednotlivých hlásek do binárního kódu) byla hláska *ch* reprezentována symbolem označujícím jinak paragraf (§).

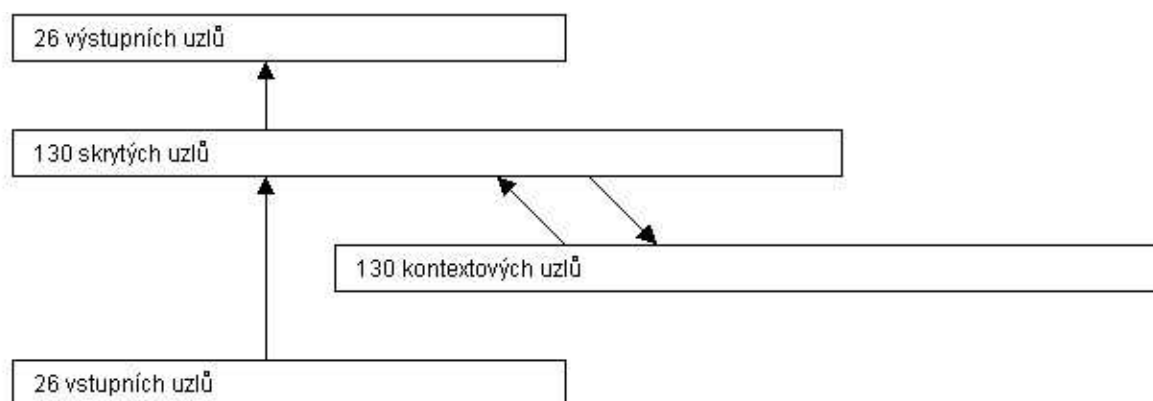
Po převedení jednotlivých hlásek do binárního reprezentačního kódu tak tréninkové prostředí obsahovalo posloupnost 2 122 (26-ti místných) vektorů jedniček a nul. Na obrázku 168 je takto zachyceno několik prvních hlásek z tréninkového korpusu:

h	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
l	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0
a	0	1	0	1	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
v	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0
a	0	1	0	1	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
s	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0
t	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0
u	0	1	0	1	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
d	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0
e	0	1	0	1	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
n	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0
t	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0
a	0	1	0	1	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
p	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0
r	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0
a	0	1	0	1	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
s	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0
e	0	1	0	1	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
v	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0
i	0	1	0	1	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
s	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0
o	0	1	0	1	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
k	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
é	0	1	0	1	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Obrázek 168: Část tréninkového korpusu v binárním reprezentačním kódu.

4.3 Funkční architektura neuronové sítě

Vzhledem k povaze řešené úlohy a použitému reprezentačnímu schématu byla v simulačním projektu použita jednoduchá rekurentní síť s 26 vstupními a výstupními uzly a se 130 skrytými a kontextovými uzly (viz obrázek 169) – více ke způsobu zpracovávání informací v tomto typu neuronové sítě viz oddíl „2.2.2.3.2 Jednoduché rekurentní sítě“.



Obrázek 169: Jednoduchá rekurentní síť použitá v simulačním projektu.

4.4 Tréninkové parametry neuronové sítě (experimentální design)

Cílem simulačního projektu bylo zjistit, jakým konkrétním způsobem (a zda vůbec) se předpokládaná delimitativní funkce prozodického vodítka v podobě stálého přízvuku na první slabice projevů v činnosti Elmanem navrženého konekcionistického modelu segmentace řeči, a ověřit tak schopnost tohoto modelu implementovat a vysvětlit delimitativní funkci stálého přízvuku na první slabice. Za tímto účelem bylo potřeba vytvořit dvě různé experimentální situace, ve kterých by jedna a ta samá neuronová síť byla vystavena dvěma různým tréninkovým korpusům, které by se lišily pouze v ne/přítomnosti prozodického vodítka majícího podobu stálého přízvuku na první slabice. Vzhledem k vysoké obtížnosti procesu kódování informace o ne/přítomnosti přízvuku bylo v průběhu simulačního experimentu manipulováno pouze s jednou nezávislou proměnnou, kterou byla ne/přítomnost přízvuku na první slabice všech (dvaačtyřiceti) slov s kořenem *koleg*. Tzn. že zatímco „kontrolní“ neuronová síť byla vystavena tréninkovému prostředí, ve kterém nebyla obsažena žádná informace o tom, kde se ve spojitě toku hlásek vyskytuje přízvuk, „experimentální“ neuronová síť měla v tréninkovém prostředí k dispozici informaci o umístění přízvuku na slabice *ko*. Z pohledu člověka by rozdíl mezi oběma tréninkovými korpusy vypadal přibližně následujícím způsobem:

visokékolegoustarésluncílegistaréstarímvisokíkolegovi...

x

*visoké**ko**legoustarésluncí**ko**legistaréstarímvisokí**ko**legovi...*

Z hlediska UNS ten samý rozdíl vypadal tak, jak je to znázorněno na obrázku 170.

Sledovány byly dvě závislé proměnné. Tou první byla globální chyba (*global error*)¹, která se měřila na konci každé tréninkové epochy, tj. pokaždé, co UNS prošla celou množinou tréninkových dat. Vzhledem k tomu, že každá ze 2122 vstupně-výstupních dvojic tréninkového korpusu byla UNS prezentována přesně 20 x (což jinak řečeno znamená, že trénování UNS probíhalo ve 20 epochách, 42 440/2 122), bylo pro obě experimentální situace k dispozici 20 různých údajů, které bylo možné mezi sebou porovnávat. Druhou závislou proměnnou byl vzorec individuálních chyb (*individual pattern error*)² pro všechna slova s kořenem *koleg*.

¹ Viz oddíl „2.3.2 Hodnotící fáze“.

² Viz „2.3.2 Hodnotící fáze“.

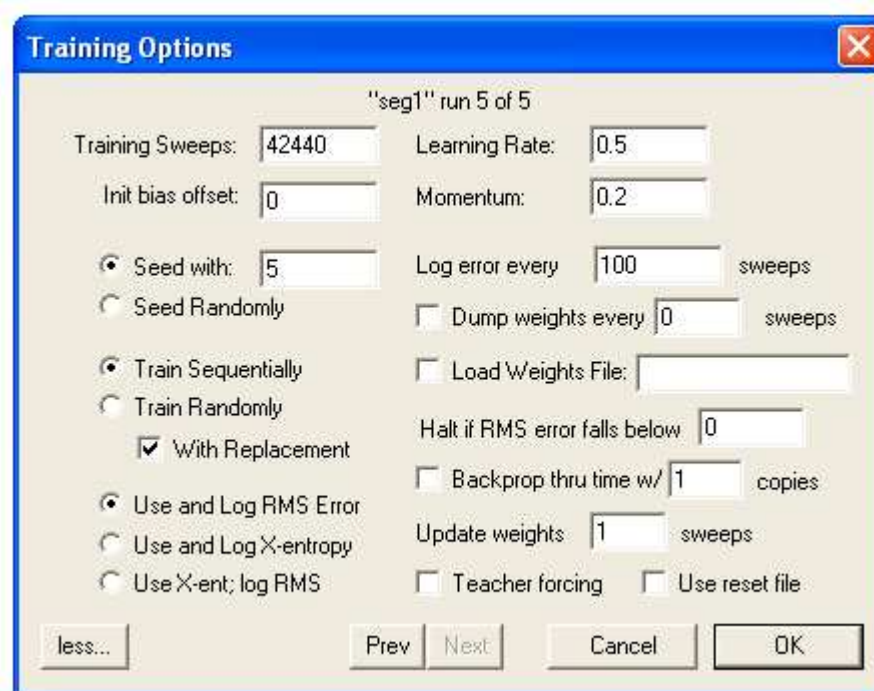
Hodnota těchto chybových údajů byla zjišťována pouze jednou a to na konci celého tréninkového procesu, tj. až po proběhnutí všech 20 tréninkových epoch.

00110000000000010000100000	v	00110000000000010000100000
01011001000100000000000000	i	01011001000100000000000000
00100000000000010000010000	s	00100000000000010000010000
01010010100100000000000000	o	01010010100100000000000000
001000000000100000000000010	k	001000000000100000000000010
01011000101000000000000000	é	01011000101000000000000000
001000000000100000000000010	k	1 0100000000010000000000010
01010010100100000000000000	o	1 0100101001000000000000000
0011000000000001000010000	l	0011000000000001000010000
01011000100100000000000000	e	01011000100100000000000000
001100000000100000000000010	g	001100000000100000000000010
01010010100100000000000000	o	01010010100100000000000000
01010011000100000000000000	u	01010011000100000000000000
00100000000000010000010000	s	00100000000000010000010000
00100000000010000000010000	t	00100000000010000000010000
01010100010100000000000000	a	01010100010100000000000000
001100000000000100010000	r	001100000000000100010000
01011000101000000000000000	é	01011000101000000000000000
00100000000000010000010000	s	00100000000000010000010000
0011000000000001000010000	l	0011000000000001000010000
01010011000100000000000000	u	01010011000100000000000000
00110000000010000000010000	n	00110000000010000000010000
001000000000000100000010000	c	001000000000000100000010000
01011001001000000000000000	í	01011001001000000000000000
001000000000100000000000010	k	1 0100000000010000000000010
01010010100100000000000000	o	1 0100101001000000000000000
0011000000000001000010000	l	0011000000000001000010000
01011000100100000000000000	e	01011000100100000000000000
001100000000100000000000010	g	001100000000100000000000010
01011001000100000000000000	i	01011001000100000000000000
00100000000000010000010000	s	00100000000000010000010000
00100000000010000000010000	t	00100000000010000000010000
01010100010100000000000000	a	01010100010100000000000000
001100000000000100010000	r	001100000000000100010000
01011000101000000000000000	é	01011000101000000000000000
00100000000000010000010000	s	00100000000000010000010000
00100000000010000000010000	t	00100000000010000000010000
01010100010100000000000000	a	01010100010100000000000000
001100000000000100010000	r	001100000000000100010000
01011001001000000000000000	í	01011001001000000000000000
00110000000001000001000000	m	00110000000001000001000000
00110000000000010000100000	v	00110000000000010000100000
01011001000100000000000000	i	01011001000100000000000000
00100000000000010000010000	s	00100000000000010000010000
01010010100100000000000000	o	01010010100100000000000000
001000000000100000000000010	k	001000000000100000000000010
01011001001000000000000000	í	01011001001000000000000000
001000000000100000000000010	k	1 0100000000010000000000010
01010010100100000000000000	o	1 0100101001000000000000000
0011000000000001000010000	l	0011000000000001000010000
01011000100100000000000000	e	01011000100100000000000000
001100000000100000000000010	g	001100000000100000000000010
01010010100100000000000000	o	01010010100100000000000000
00110000000000010000100000	v	00110000000000010000100000
01011001000100000000000000	i	01011001000100000000000000

Obrázek 170: Porovnání části tréninkového prostředí umělé neuronové sítě ve dvou různých experimentálních podmínkách (nalevo bez přízvuku, napravo s přízvukem).

Stejně jako v tradičním i v simulačním experimentu je potřeba zajistit, aby do zkoumání vztahů a souvislostí mezi nezávislou a závislou proměnnou nemohly zasahovat nejrůznější nežádoucí proměnné. Výhodou konekcionistického modelování je to, že experimentátor má absolutní kontrolu jak nad parametry (tréninkového) prostředí a charakteristikami zkoumaných subjektů, tak i nad způsobem jejich vzájemné interakce. Z předcházejícího výkladu o povaze tréninkového prostředí ve dvou různých experimentálních podmínkách by mělo být zřejmé, že jediný rozdíl mezi těmito dvěma typy prostředí spočíval v ne/přítomnosti informace o místě přízvuku na slabice *ko* ve všech slovech s kořenem *koleg*. Tzn. že ať už bude zjištěn jakýkoli rozdíl mezi výkonem UNS v těchto dvou různých experimentálních podmínkách, můžeme se 100 % jistotou tvrdit, že příčinou tohoto rozdílu může být z charakteristik (tréninkového) prostředí jediné ne/přítomnost prozodického vodítka. Abychom však něco takového mohli tvrdit, musíme mít jistotu, že UNS vystavená dvěma různým experimentálním podmínkám s různě nastavenou hladinou nezávislé proměnné bude mít v obou těchto situacích přesně tytéž „vnitřní“ charakteristiky. Jedině tak můžeme později vyloučit tu možnost, že by rozdíly ve výkonech UNS byly způsobeny odlišnými „vnitřními dispozicemi“ UNS. S tímto úzce souvisí to, na jakém konkrétním místě UNS započne svou „pout“ chybovým prostorem, a způsob, jakým UNS bude data načítat ze svého (tréninkového) prostředí. Konstantnost všech těchto podmínek napříč různými experimentálními situacemi lze u UNS jednoduše zajistit nastavením několika jejích základních tréninkových parametrů jako je míra učení, moment hybnosti, použitá aktivační funkce, počet tréninkových epoch, způsob prezentace vstupních dat (náhodný/sekvenční), seed - parametr, který je součástí algoritmu produkujícího pseudonáhodná čísla, která jsou simulátorem používána při generování výchozí váhové matice - a nakonec rozmezí hodnot, ve kterém se může pohybovat výchozí náhodné nastavení vah spojuj mezi uzly. Konkrétní hodnoty všech těchto tréninkových parametrů¹ (které byly společné pro obě použité UNS) jsou zachyceny na obrázku 171, kde je zobrazeno příslušné dialogové okno ze simulačního softwaru Tlearn, s pomocí kterého byl konekcionistický model vytvořen.

¹ S výjimkou použité aktivační funkce a rozmezí hodnot, v němž se pohybovalo výchozí náhodné nastavení vah. Oba tyto údaje se nachází v souboru specifikujícím architekturu neuronové sítě a aktivační funkci jednotlivých uzlů. V případě zde prezentovaného konekcionistického modelu byla použita běžná logistická aktivační funkce v kombinaci s lineární aktivační funkcí (to pro zajištění schopnosti kontextové vrstvy plnit svou „paměťovou“ funkci) a rozmezí $\pm 0,1$. K trénování sítě byl použit běžný učící algoritmus zpětného šíření (*backpropagation learning algorithm*).



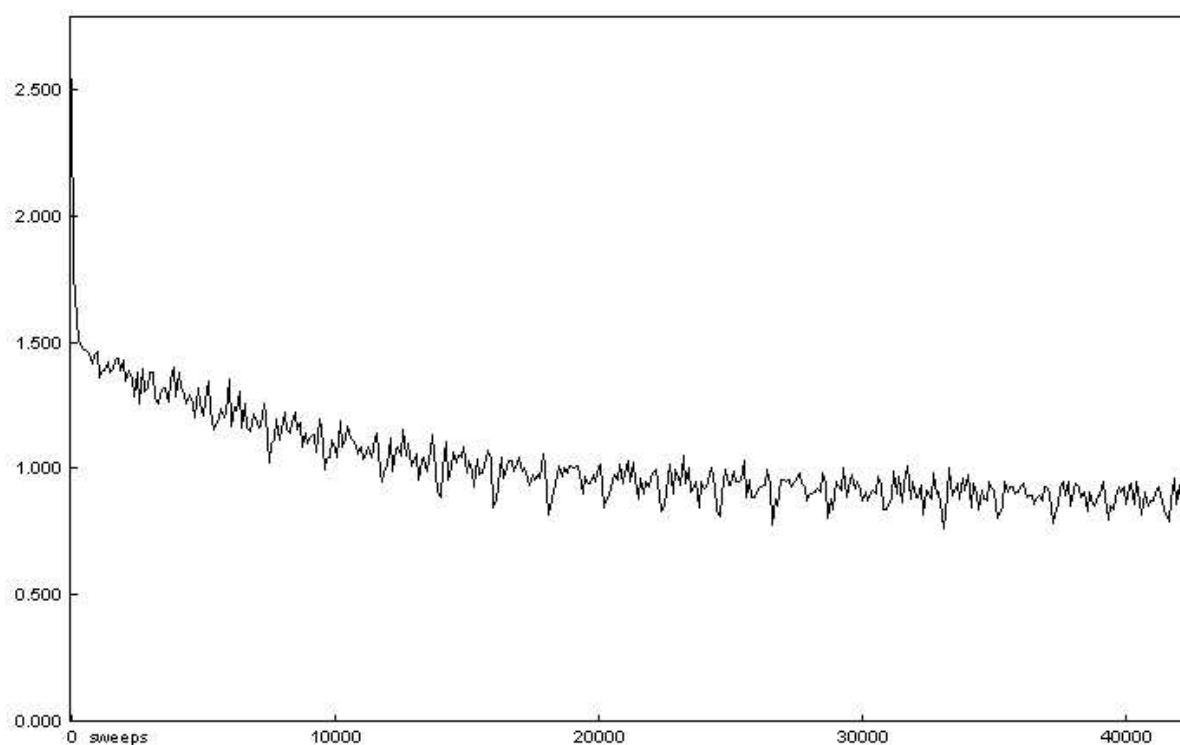
Obrázek 171: Dialogové okno ze simulačního softwaru Tlearn, ve kterém jsou zaneseny všechny důležité tréninkové parametry, které byly použity při tréninku neuronových sítí.

4.5 Některé předběžné výsledky

S připraveným tréninkovým prostředím a nastavenými tréninkovými parametry bylo možné přistoupit k samotné simulaci. To, že pořadí slov v tréninkovém prostředí bylo zcela náhodné (tzn. že například slovo *student* mohlo v tréninkovém korpusu následovat po třech různých slovech) a že jazykový vstup tvořily (de facto) kořeny slov spojené s různými pádovými koncovkami, znamenalo, že pořadí jednotlivých hlásek bylo jen částečně předpověditelné. Na základě této skutečnosti tak bylo možné již dopředu předvídat, že globální chyba, která je celkovým ukazatelem chybovosti UNS napříč celým souborem vstupně-výstupních dvojic, se poměrně záhy po započetí procesu učení ustálí na relativně vysokých hodnotách, které se budou nacházet daleko od „ideální“ nulové hladiny chybovosti. Tento předpoklad se plně potvrdil, a to v obou experimentálních podmínkách. V grafech 26 a 27 jsou zobrazeny vývojové křivky chybového ukazatele UNS ve dvou experimentálních podmínkách, na kterých je dobře vidět, jak se globální chyba v obou dvou případech nejdříve prudce snížila z původních 2,5 na 1,5, poté již mnohem pozvolněji klesla na hodnotu 1 a nakonec se (v rámci námi sledovaného úseku 20 epoch) ustálila okolo hodnoty 0,9.¹ Takto

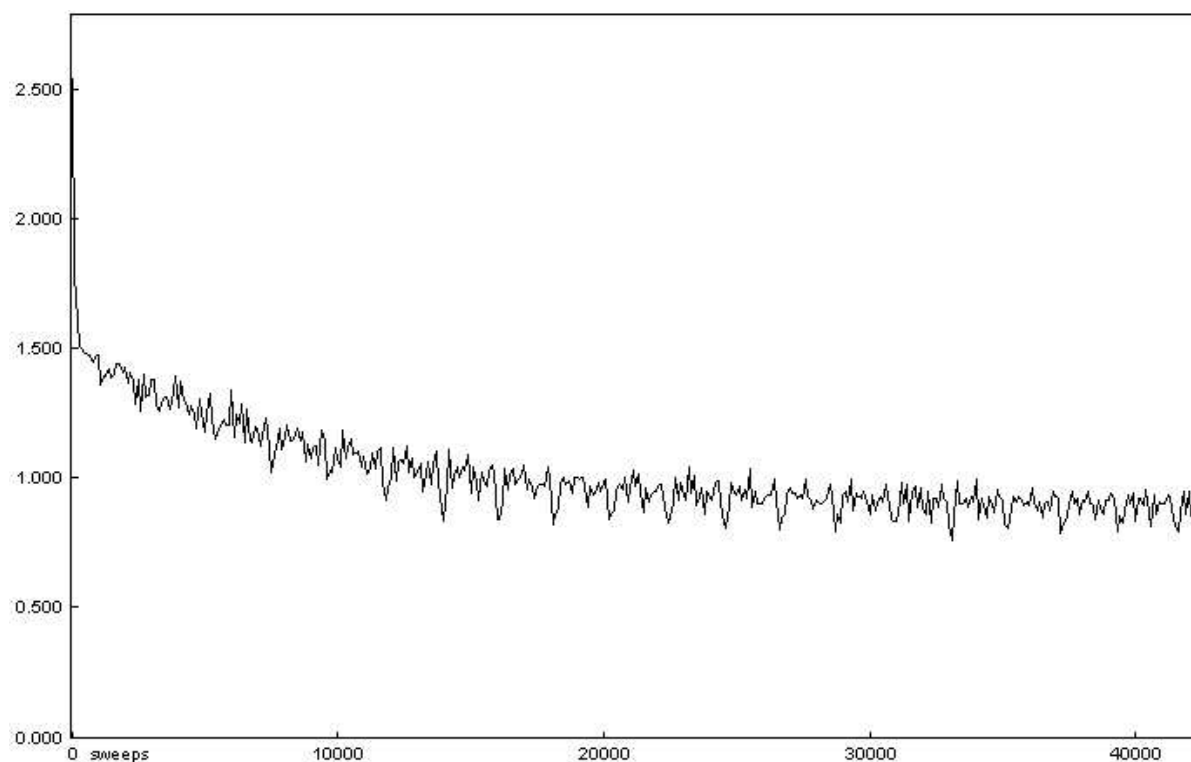
¹ U obou UNS měla globální chyba tendenci se neustále, zato však velmi pomalu a nepatrně zmenšovat. Jenom pro představu, jakým tempem k této změně docházelo, uvádím hodnotu globální chyby na konci 20. a 100.

vysoká hladina chybovosti je nevyhnutelným důsledkem samotné povahy statistické struktury tréninkového prostředí, ve kterém je pořadí hlásek do značné míry náhodné. Chybovost by samozřejmě bylo možné snížit zvýšením počtu skrytých uzlů, takže UNS by nakonec byla schopná se celý tréninkový korpus naučit mechanicky nazpaměť, to by však bylo na úkor její schopnosti generalizace.



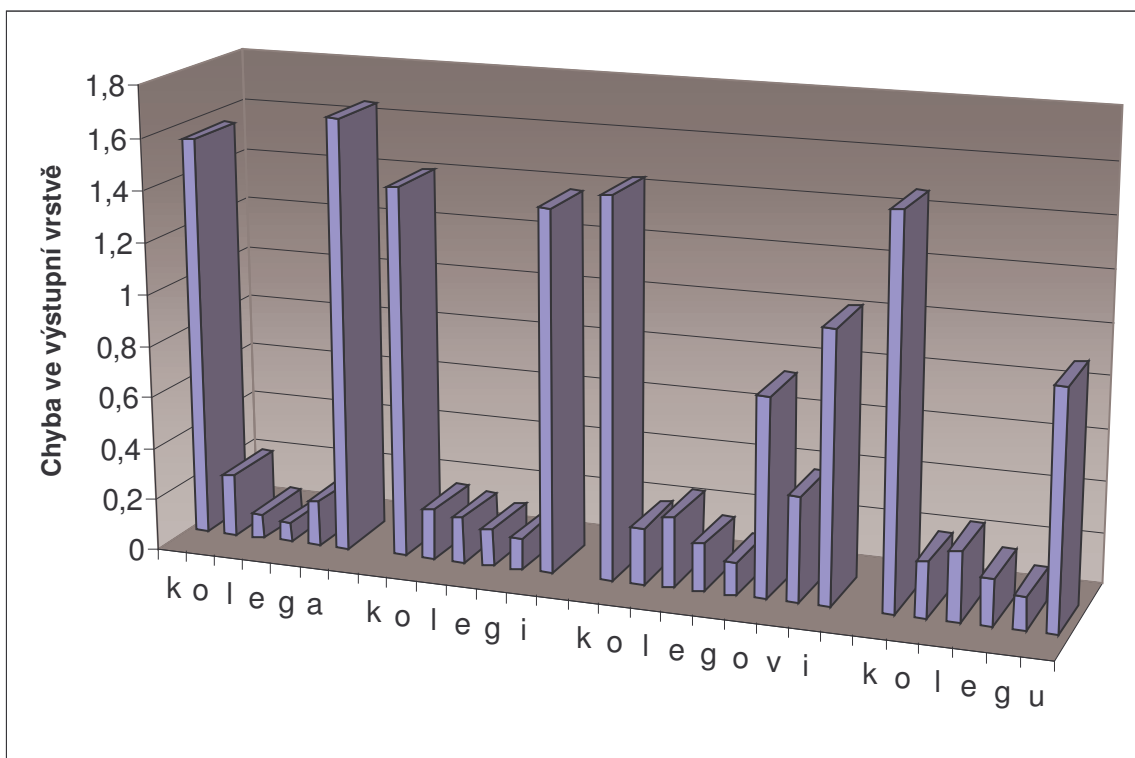
Graf 26: Globální chyba umělé neuronové sítě, která ve svém tréninkovém prostředí neměla k dispozici prozodická vodítka. Nejvyšší míra chybovosti je na samotném začátku, kdy UNS začíná s náhodně nastavenými vahami jednotlivých spojů. S jejich postupnými úpravami se výkon UNS zlepšuje, nejdříve prudce a potom již jen velice pozvolna. Každá ze 2 122 vstupně-výstupních dvojic tréninkového korpusu byla UNS prezentována 20 x (42 440/2 122). (vytvořeno s pomocí simulačního softwaru Tlearn)

epochy, která činila 0,897926, resp. 0,784864 (údaje platí pro UNS, která neměla k dispozici prozodická vodítka).

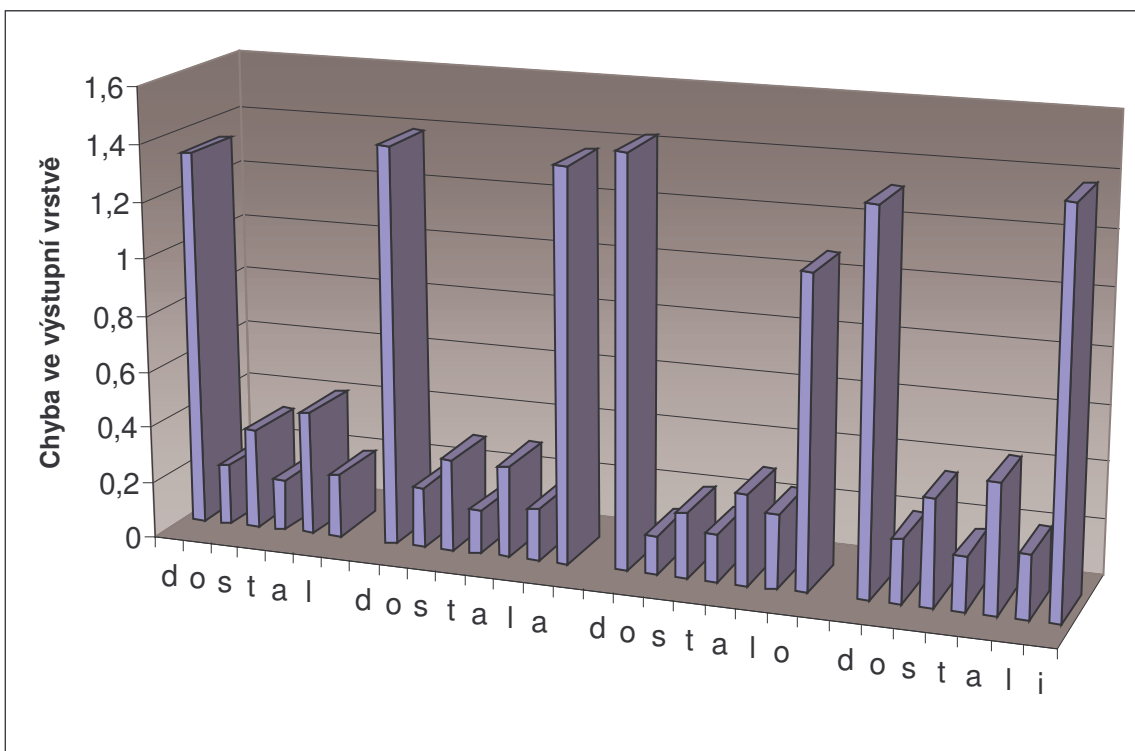


Graf 27: Globální chyba umělé neuronové sítě, která ve svém tréninkovém prostředí měla k dispozici informaci o umístění stálého přízvuku na první slabice slov s kořenem „koleg“. Stejně jako u předchozí UNS, i zde každá ze 2 122 vstupně-výstupních dvojic tréninkového korpusu byla UNS prezentována 20 x (42 440/2 122). (vytvořeno s pomocí simulačního softwaru Tlearn)

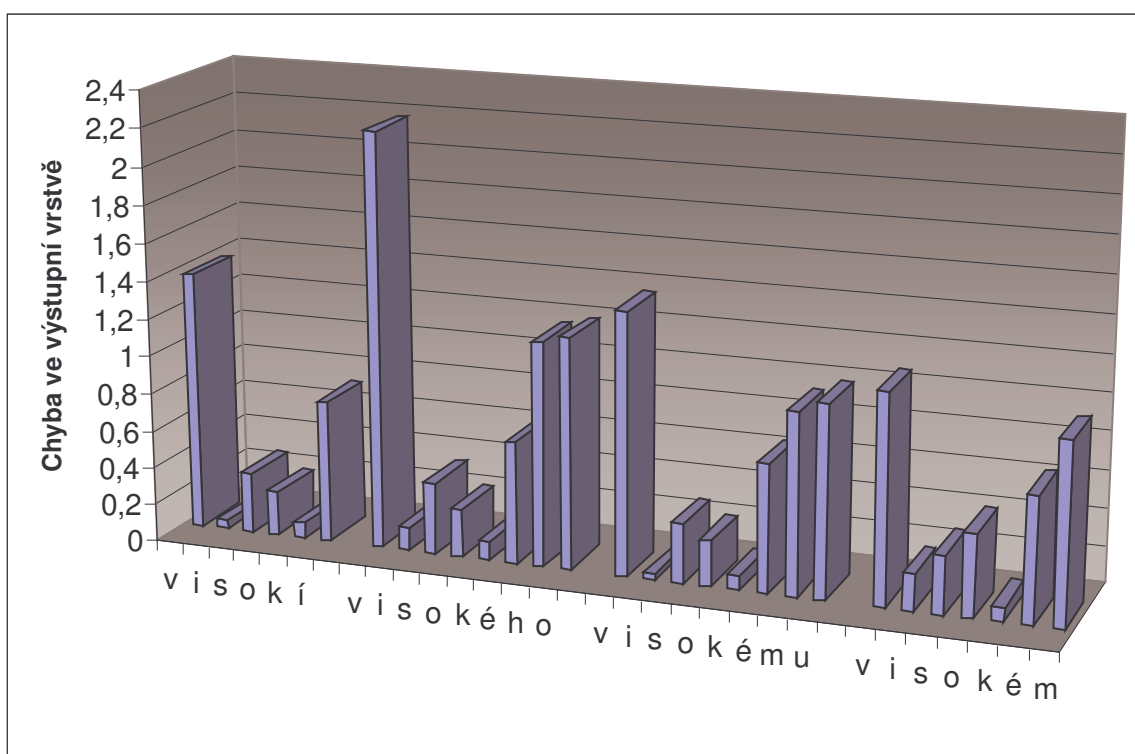
Na základě znalosti statistické struktury tréninkového prostředí bylo rovněž možné již dopředu očekávat, že na konci celého tréninkového procesu budou mít vzorce individuálních chyb pro jednotlivá slova tvar (více či méně pravidelného) písmene U: To, že první hláska každého slova mohla následovat po několika různých hláskách (tvořících konec různých slov, která byla součástí tréninkového korpusu) a že po kořenu slova mohly následovat různé pádové koncovky, by mělo mít u UNS za následek vysokou míru chybovosti (tj. jakési její „vnitřní nejistoty“) při predikci všech hlásek, které se nacházejí na začátku slova, resp. bezprostředně za jeho kořenem. Jinak řečeno, podmíněná pravděpodobnost výskytu těchto hlásek za předpokladu výskytu všech předcházejících hlásek by měla být relativně nízká - ve srovnání s kořeny slov, které se nemění a zůstávají stabilní napříč celým tréninkovým korpusem, takže UNS by při jejich anticipaci (resp. při anticipaci hlásek, ze kterých se skládají) měla vykazovat relativně nízkou míru chybovosti („nejistoty“), která v „obklíčení“ dvou vysokých hodnot tohoto ukazatele dává vzniknout vzorci individuálních chyb ve tvaru písmene U. Grafické znázornění několika vzorců individuálních chyb pro několik námtkou vybraných slov se zdá tuto domněnku potvrzovat – viz grafy 28-30.



Graf 28: Vzorec individuálních chyb pro čtyři různá slova s kořenem „koleg“ (uvedené údaje platí pro UNS, která neměla k dispozici prozodická vodítka).



Graf 29: Vzorec individuálních chyb pro čtyři různá slova s kořenem „dostal“ (uvedené údaje platí pro UNS, která neměla k dispozici prozodická vodítka).



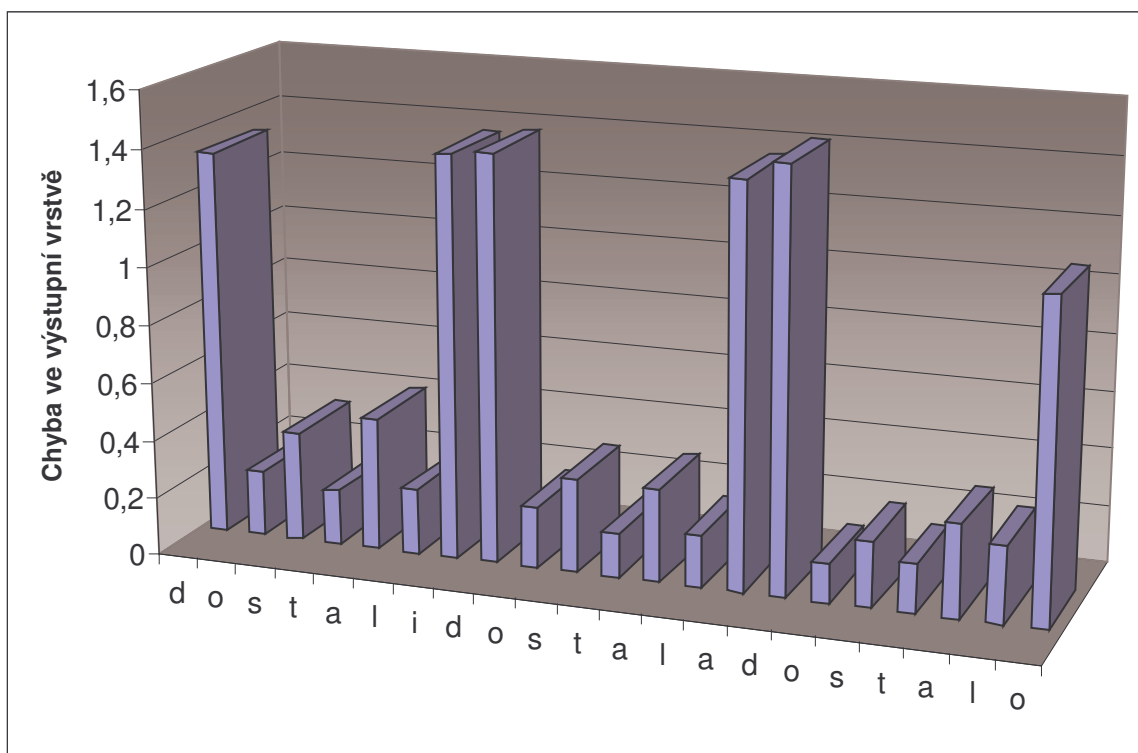
Graf 30: Vzorec individuálních chyb pro čtyři různá slova s kořenem „visok“ (uvedené údaje platí pro UNS, která neměla k dispozici prozodická vodítka).

Tyto výsledky na jedné straně potvrzují Elmanovo původní zjištění - totiž že „stěna“ jazykového vstupu není tak „hladká“, jak by se mohlo „na první poslech“ zdát, ale že se v ní nachází řada trhlinek a spár v podobě statistických zákonitostí ve výskytu jednotlivých hlásek, které jednoduchá rekurentní síť (jakožto model zpětnovazebně zapojených neuronálních struktur nacházejících se v lidském mozku) anticipující svůj příští vstup dokáže objevit – a na straně druhé ho rozšiřují o poznatek, že stejný mechanismus, který lze použít k detekci hranic mezi slovy, je možné použít také k lokalizaci hranic mezi kořenem a gramatickou koncovkou v rámci jednoho slova. Tento mechanismus tak může hrát důležitou roli při osvojování jazykové flexe (skloňování a časování), která na straně mluvčího předpokládá určitou schopnost odlišit od sebe (třebaže jen nevědomě) obsahové a gramatické morfémy.

4.5.1 Nová definice hranice slov

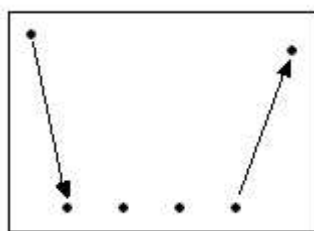
Aplikace Elmanova modelu na flektivní jazyk jakým je čeština navíc odhalila, že je potřeba mírně pozměnit původní pojetí hranice slova, kterou Elman chápal jako takové místo ve spojitém toku řeči, kde si je UNS velmi málo „jistá“ ohledně toho, jaká další hláska bude následovat, tedy jako místo s velmi nízkou podmíněnou pravděpodobností výskytu dané

Graf 31: Vzorec individuálních chyb pro tři různá slova s kořenem „dostal“ (uvedené údaje platí pro UNS, která neměla k dispozici prozodická vodítka).



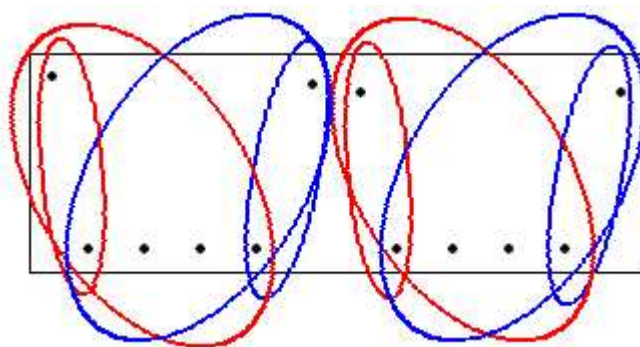
Graf 31: Vzorec individuálních chyb pro tři různá slova s kořenem „dostal“ (uvedené údaje platí pro UNS, která neměla k dispozici prozodická vodítka).

Tento problém lze poměrně jednoduše vyřešit tím, že se nově definuje **hranice slova** (a gramatické koncovky) **jako místo výrazného kontrastu (rozdílu) mezi podmíněnou pravděpodobností výskytu dvou anebo více sousedících hlásek**, přičemž v případě hranice mezi slovy by se jednalo o výrazné zvýšení podmíněné pravděpodobnosti výskytu následujících hlásek a v případě hranice mezi kořenem slova a gramatickou koncovkou naopak o její výrazné snížení - viz obrázek 172, na kterém je schematicky zachycen odpovídající tvar vzorce individuálních chyb ve výstupní vrstvě UNS zpracovávající



Obrázek 172: Schematické znázornění nové definice hranice slov (a gramatické koncovky).

posloupnost šesti hlásek, kde náhlý pokles chybovosti („nejistoty“), resp. její prudký nárůst odpovídá začátku dalšího slova, resp. hranici mezi kořenem slova a gramatickou koncovkou. Tímto způsobem pojímaná hranice slov (a gramatických koncovek) již umožňuje mezi sebou jasně rozlišit začátek slova a začátek gramatické koncovky, neboť nyní je každé z těchto dvou částí slova přiřazen jeden specifický chybový vzorec - viz obrázek 173, na kterém je schematicky znázorněn vzorec individuálních chyb pro dvě těsně za sebou vyřčená slova; minimální (nebo nedostatečný) rozdíl mezi podmíněnou pravděpodobností výskytu dvou hlásek tvořících začátek slova a gramatickou koncovku znemožňuje, aby (na základě nově



Obrázek 173: Schematické znázornění vzorce individuálních chyb pro dvě těsně za sebou vyřčená slova. Červenými ovály jsou zvýrazněny „pozitivní“ rozdíly mezi podmíněnou pravděpodobností výskytu dvou nebo více sousedících hlásek, které jsou příznakem nebo markerem hranice mezi dvěma různými slovy. Modrými ovály jsou naopak zvýrazněny „negativní“ rozdíly, které signalizují přítomnost hranice mezi kořenem slova a gramatickou koncovkou.

přijaté definice hranice slov) došlo k záměně těchto dvou různých částí dvou různých slov.

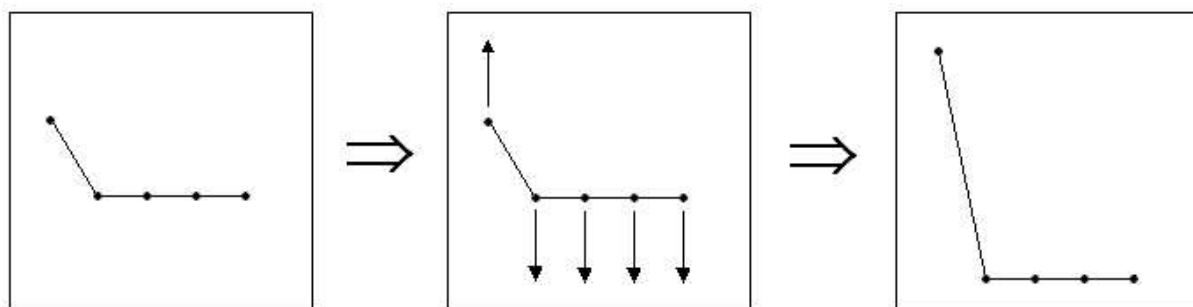
Tato nová definice hranice slov má vliv rovněž na hledání odpovědi na hlavní otázku tohoto simulačního experimentu, totiž jakým způsobem se projeví vliv prozodického vodítka v činnosti UNS a zda Elmanem navržený konekcionistický model segmentace řeči dokáže implementovat a vysvětlit delimitativní funkci stálého přízvuku na první slabice.

4.5.2 Formulace hypotéz

Na první pohled je poměrně obtížné si představit, jak by Elmanův konekcionistický model segmentace řeči mohl implementovat delimitativní funkci stálého přízvuku, když víme, že informace o hranici slov je v tomto modelu de facto pouze vedlejším produktem neznalosti či „nejistoty“ UNS tváří v tvář úloze anticipace svého příštího vstupu: Jestliže hranicí slova je takové místo ve spojitém toku řeči, kde je jen velmi nízká podmíněná pravděpodobnost výskytu další hlásky, potom to znamená, že delimitativní funkce prozodického vodítka se v Elmanově modelu může projevit jediné tak, že se tato podmíněná pravděpodobnost ještě sníží, tedy že se zvýší „nejistota“ UNS ohledně jejího dalšího vstupu. V případě, že by mělo dojít k přesnému opaku, tedy ke zvýšení podmíněné pravděpodobnosti a ke snížení „nejistoty“ UNS ohledně svého příštího vstupu, potom by to mělo za následek jen ztrátu vlastní informace o tom, kde se hranice slov nachází. Tento závěr je trochu v rozporu s intuitivním pocitem, že prozodické vodítko v podobě stálého přízvuku by mělo nejistotu ohledně umístění hranice slov naopak spíše snižovat. Přesto je to jediný rozumný závěr, který neúprosně a logicky vyplývá ze způsobu, jakým je v Elmanově modelu konceptualizován pojem hranice slova.

Poněkud odlišnou podobu získává delimitativní efekt stálého přízvuku v kontextu nové definice hranice slov, kde je hodnota podmíněné pravděpodobnosti výskytu hlásky tvořící počátek slova pouze jednou ze dvou „proměnných“, které jsou součástí „rovnice“ lokalizující hranici slov. Podle nové definice je začátkem dalšího slova místo výrazného rozdílu mezi podmíněnou pravděpodobností výskytu dvou anebo více sousedících hlásek, tzn. že delimitativní funkce stálého přízvuku by se měla projevit především zvětšením rozdílu (kontrastu) mezi podmíněnou pravděpodobností výskytu hlásky, která tvoří začátek nového slova, a podmíněnou pravděpodobností výskytu jedné nebo více hlásek, které ji bezprostředně následují. Tento efekt přitom může být způsoben dvěma mechanismy:

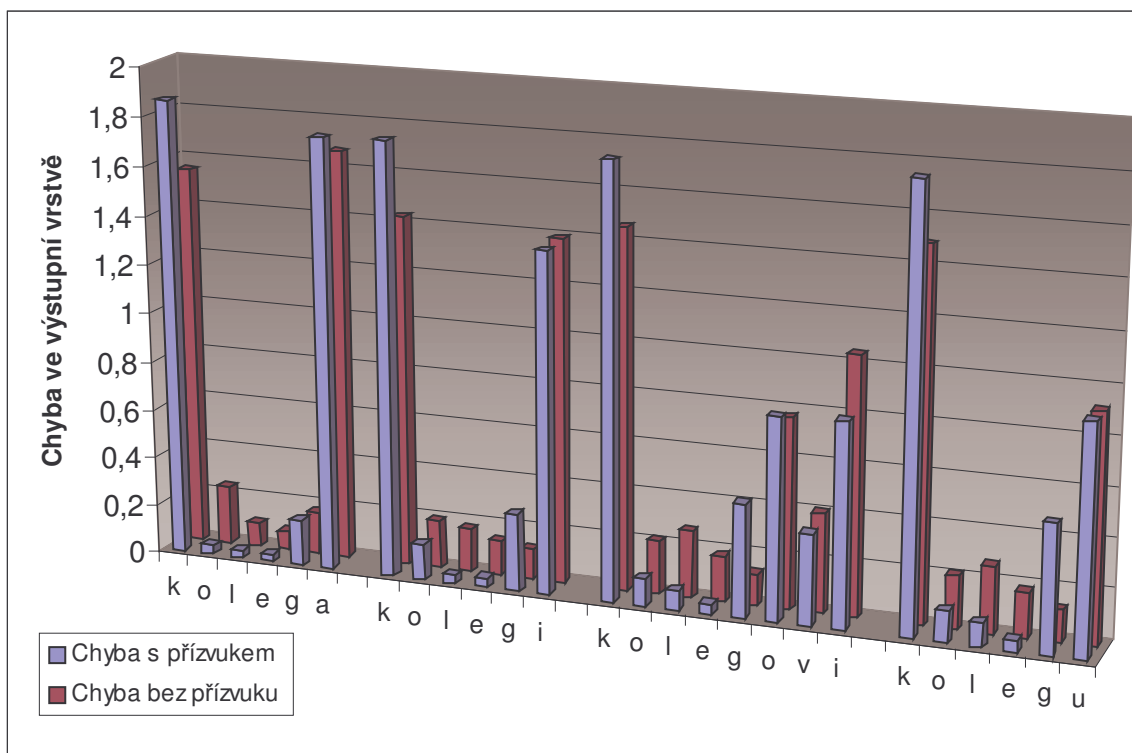
- snížením podmíněné pravděpodobnosti výskytu hlásky tvořící začátek slova nebo naopak
- zvýšením podmíněné pravděpodobnosti výskytu jedné nebo více hlásek, které po první hlásce nového slova bezprostředně následují (viz obrázek 174).



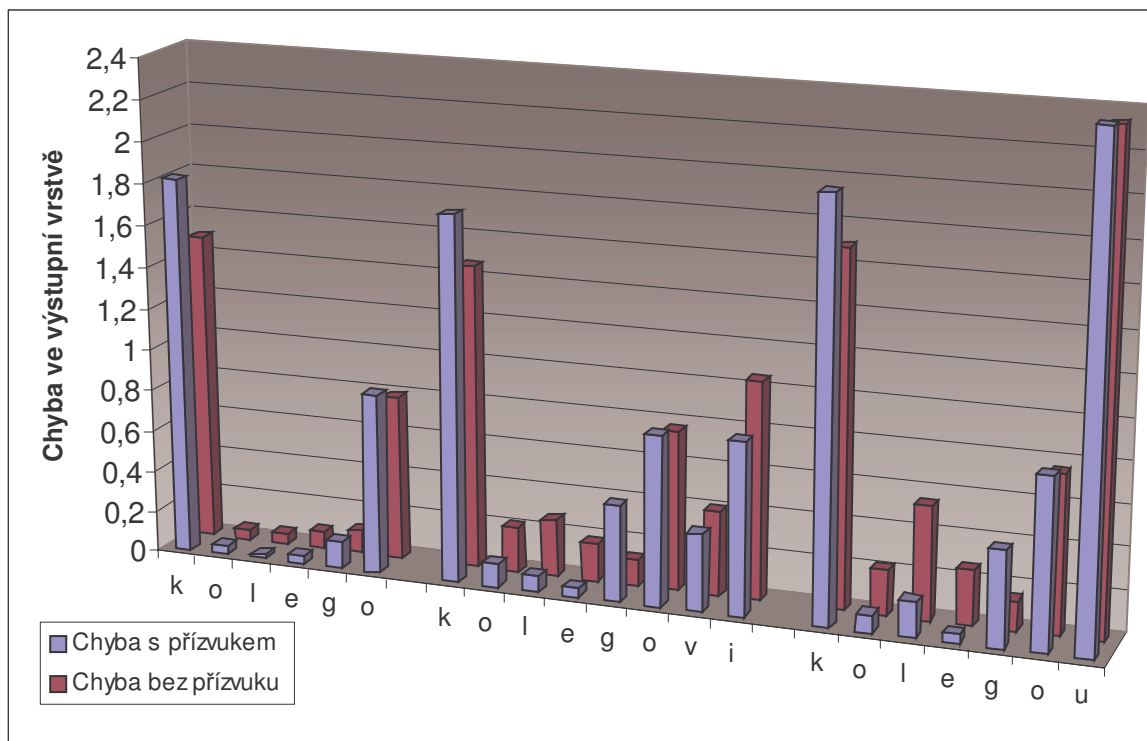
Obrázek 174: Schematické znázornění předpokládaného vlivu přítomnosti prozodického vodítka v tréninkovém prostředí UNS na vzorec individuálních chyb pro jedno slovo. V případě podmíněných pravděpodobností výskytu jednotlivých hlásek by odpovídající schéma bylo pouze zrcadlově převrácené. Na obrázku jsou zachyceny hlásky tvořící kořen slova.

Je přitom vysoce pravděpodobné, že kombinací těchto dvou mechanismů dojde k tomu, že se jejich vliv na vzorec individuálních chyb vyruší, takže na úrovni globální chyby, která je celkovým ukazatelem chybovosti UNS napříč celým souborem vstupně-výstupních dvojic, by se výkon UNS vystavené dvěma různým tréninkovým prostředím (s prozodickým vodítkem a bez něj) neměl statisticky významně lišit - ve smyslu větší nebo menší celkové chybovosti UNS, jejíž tréninkový korpus obsahoval prozodickou informaci v podobě stálého přízvuku na první slabice.

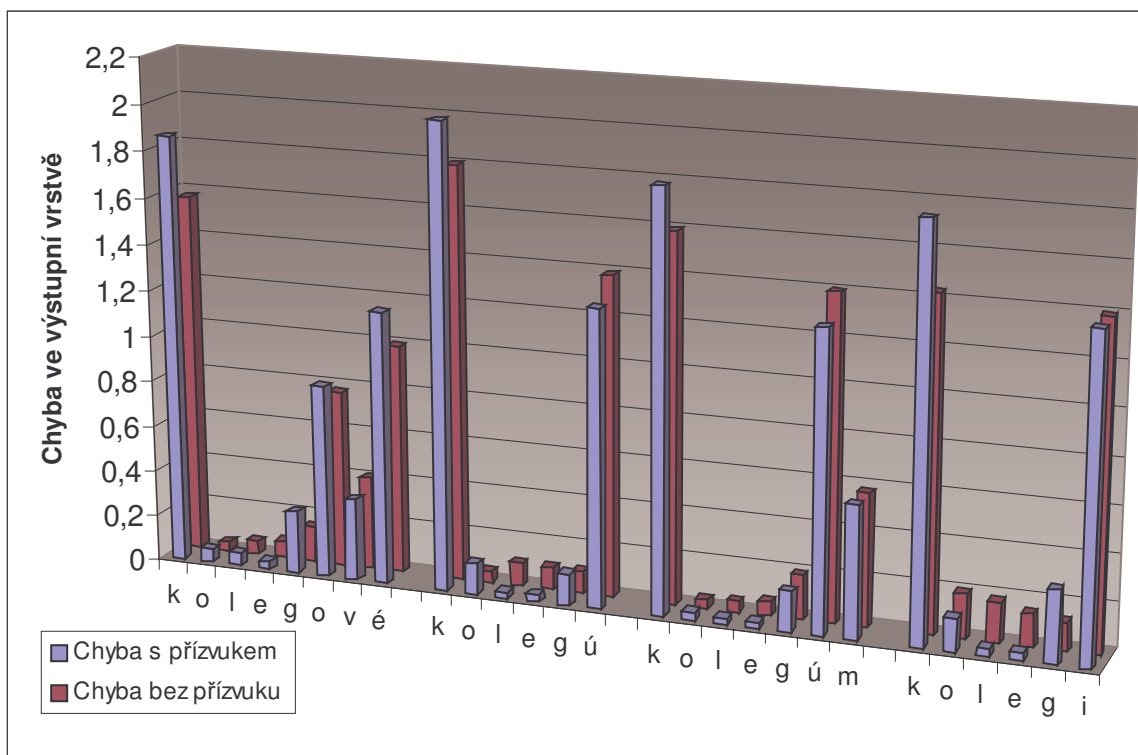
Plausibilitu některých z výše uvedených domněnek o povaze vlivu prozodického vodítka na chování UNS se zdá podporovat několik vzorců individuálních chyb pro několik namátkou vybraných slov s kořenem *koleg*, které pokrývají všechny jejich možné gramatické role (ve smyslu gramatických pádů a čísel). Tyto vzorce individuálních chyb jsou pro větší názornost zobrazeny v podobě 3-D sloupcových grafů (32-35), ze kterých je velice dobře patrné, že poté, co je do tréninkového prostředí UNS přidána informace o umístění stálého přízvuku na první slabice slov s kořenem *koleg*, zde existuje tendence ke zvyšování „nejistoty“ UNS při predikci první hlásky nového slova a naopak ke snižování chybovosti UNS při anticipaci (zejména prvních tří) hlásek bezprostředně následujících po této první hlásce.



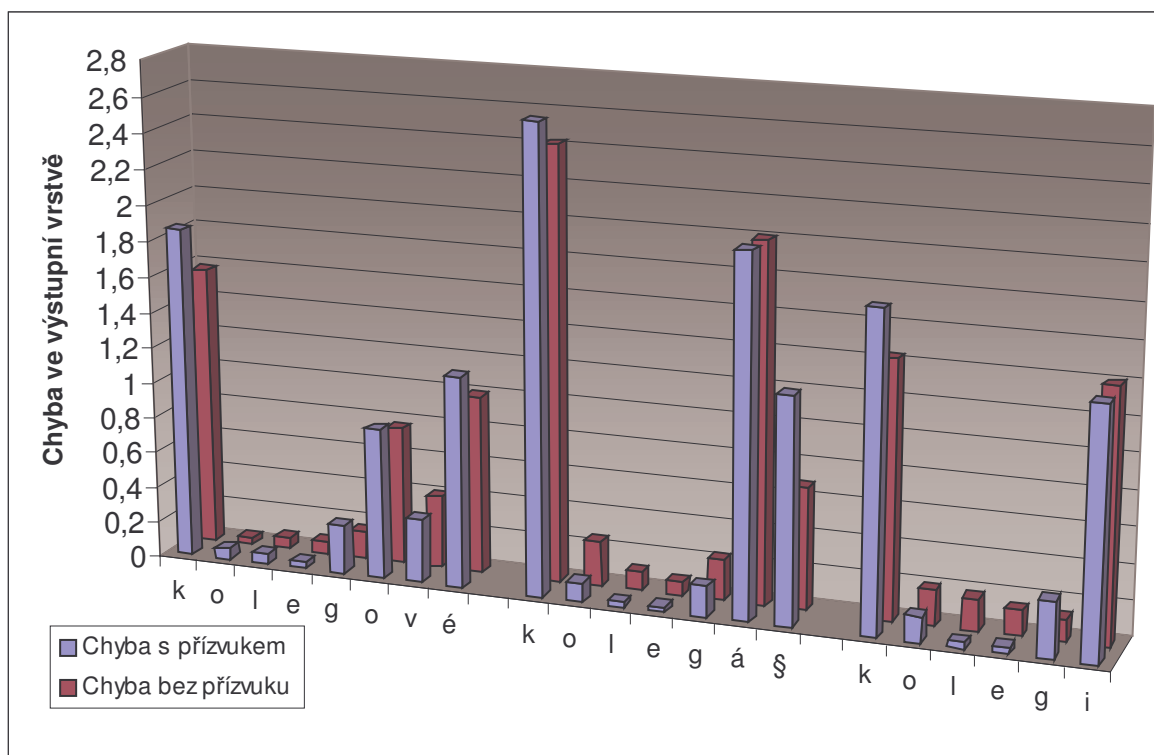
Graf 32: Vzorec individuálních chyb pro čtyři různá slova s kořenem „koleg“ ve dvou různých experimentálních podmínkách, kdy UNS buď měla, nebo naopak neměla k dispozici informaci o umístění přízvuku.



Graf 33: Vzorec individuálních chyb pro tři různá slova s kořenem „koleg“ ve dvou různých experimentálních podmínkách, kdy UNS buď měla, nebo naopak neměla k dispozici informaci o umístění přízvuku.



Graf 34: Vzorec individuálních chyb pro čtyři různá slova s kořenem „koleg“ ve dvou různých experimentálních podmínkách, kdy UNS buď měla, nebo naopak neměla k dispozici informaci o umístění přízvuku.



Graf 35: Vzorec individuálních chyb pro čtyři různá slova s kořenem „koleg“ ve dvou různých experimentálních podmínkách, kdy UNS buď měla, nebo naopak neměla k dispozici informaci o umístění přízvuku.

Na základě výše uvedených úvah a za účelem rigoróznějšího ověření s nimi spojených domněnek o povaze vlivu prozodického vodítka na činnost UNS můžeme formulovat čtyři následující nulové (a jim odpovídající alternativní) hypotézy:

Hypotéza 1

H₀: Umělá neuronová síť exponovaná tréninkovému prostředí, které neobsahuje žádnou informaci o umístění stálého přízvuku, a umělá neuronová síť exponovaná tréninkovému prostředí, které obsahuje informaci o umístění stálého přízvuku na první slabice všech slov s kořenem „koleg“, se neliší v rozložení rozdílu mezi chybou při predikci hlásky „k“ na začátku všech slov s kořenem „koleg“ a průměrnou chybou při predikci hlásek „o“, „l“, „e“ a „g“, které jsou součástí kořene těchto slov a které bezprostředně následují po jejich první hlásce. (H_A: Umělá neuronová síť exponovaná tréninkovému prostředí, které obsahuje informaci o umístění stálého přízvuku na první slabice všech slov s kořenem koleg, vykazuje větší rozdíl mezi chybou při predikci hlásky k na začátku všech slov s kořenem koleg a průměrnou chybou při predikci hlásek, o, l, e a g, které jsou součástí kořene těchto slov a které bezprostředně následují po jejich první hlásce, než umělá neuronová síť exponovaná tréninkovému prostředí, které neobsahuje žádnou informaci o umístění stálého přízvuku.)

Hypotéza 2

H₀: Umělá neuronová síť exponovaná tréninkovému prostředí, které neobsahuje žádnou informaci o umístění stálého přízvuku, a umělá neuronová síť exponovaná tréninkovému prostředí, které obsahuje informaci o umístění stálého přízvuku na první slabice všech slov s kořenem „koleg“, se neliší v rozložení chyby při predikci hlásky „k“ na začátku všech slov s kořenem „koleg“. (H_A: Umělá neuronová síť exponovaná tréninkovému prostředí, které obsahuje informaci o umístění stálého přízvuku na první slabice všech slov s kořenem koleg, vykazuje větší chybu při predikci hlásky k na začátku všech slov s kořenem koleg než umělá neuronová síť exponovaná tréninkovému prostředí, které neobsahuje žádnou informaci o umístění stálého přízvuku.)

Hypotéza 3

H₀: Umělá neuronová síť exponovaná tréninkovému prostředí, které neobsahuje žádnou informaci o umístění stálého přízvuku, a umělá neuronová síť exponovaná tréninkovému prostředí, které obsahuje informaci o umístění stálého přízvuku na první slabice všech slov s kořenem „koleg“, se neliší v rozložení průměrné chyby při predikci hlásek „o“, „l“, „e“,

„g“, které jsou součástí kořene všech slov s kořenem „koleg“ a které bezprostředně následují po první hlásce těchto slov. (H_A : Umělá neuronová síť exponovaná tréninkovému prostředí, které obsahuje informaci o umístění stálého přízvuku na první slabice všech slov s kořenem „koleg“, vykazuje menší průměrnou chybu při predikci hlásek „o“, „l“, „e“, „g“, které jsou součástí kořene všech slov s kořenem „koleg“ a které bezprostředně následují po první hlásce těchto slov, než umělá neuronová síť exponovaná tréninkovému prostředí, které neobsahuje žádnou informaci o umístění stálého přízvuku.)

Hypotéza 4

H_0 : Umělá neuronová síť exponovaná tréninkovému prostředí, které neobsahuje žádnou informaci o umístění stálého přízvuku, a umělá neuronová síť exponovaná tréninkovému prostředí, které obsahuje informaci o umístění stálého přízvuku na první slabice všech slov s kořenem „koleg“, se neliší v rozložení globální chyby měřené v průběhu 20 tréninkových epoch. (H_A : Umělá neuronová síť exponovaná tréninkovému prostředí, které obsahuje informaci o umístění stálého přízvuku na první slabice všech slov s kořenem „koleg“, vykazuje v průběhu 20 tréninkových epoch vyšší nebo nižší globální chybu než umělá neuronová síť exponovaná tréninkovému prostředí, které neobsahuje žádnou informaci o umístění stálého přízvuku.)

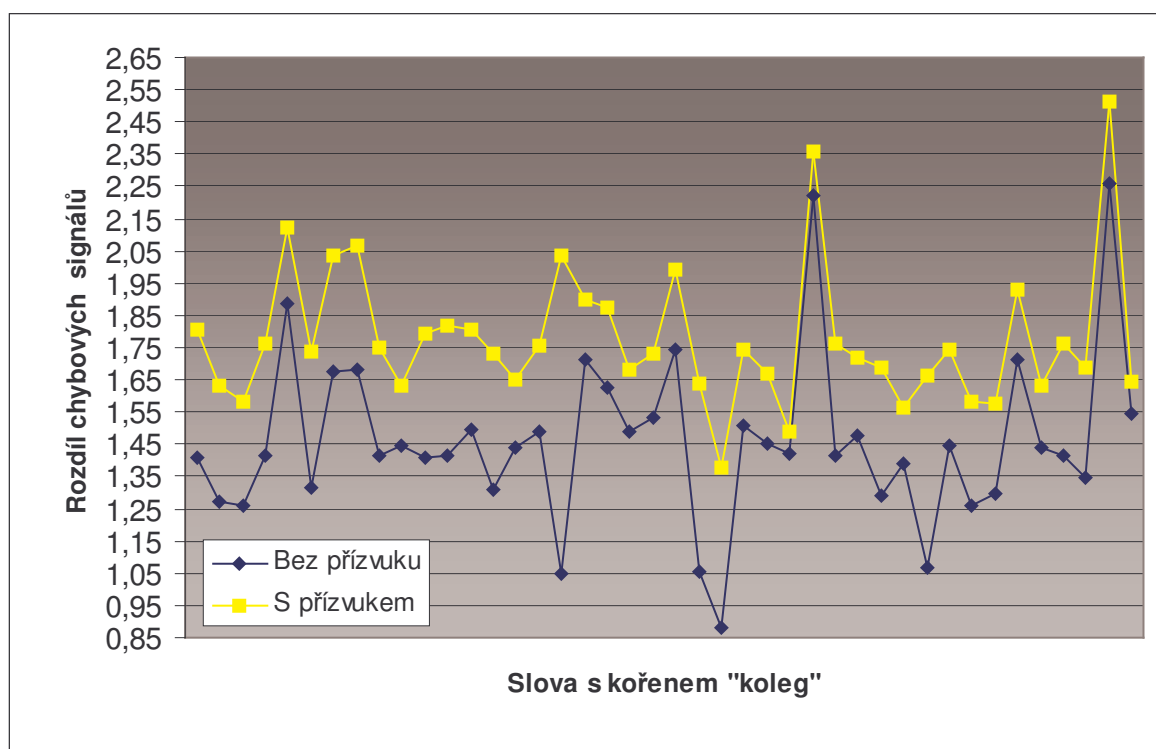
První tři hypotézy byly testovány pomocí Wilcoxonova neparametrického pořadového testu pro párované hodnoty, vždy s jednostrannou signifikancí a s požadovanou hladinou významnosti $\alpha = 0,01$. Čtvrtá hypotéza pak byla testována pomocí Friedmanova testu s dvoustrannou signifikancí a s požadovanou hladinou významnosti $\alpha = 0,01$. Ke statistické analýze dat získaných v průběhu simulačního experimentu byl použit statistický software XLSTAT.

4.6 Výsledky a diskuse

V rámci testování Hypotézy 1 nás zajímá, zda se přítomnost prozodického vodítka v tréninkovém korpusu nějak projeví ve velikosti rozdílu mezi chybou při predikci hlásky *k* na začátku všech (dvaačtyřiceti) slov s kořenem *koleg* a průměrnou chybou při predikci hlásek, *o*, *l*, *e* a *g*, které jsou součástí kořene těchto slov a které bezprostředně následují po jejich první hlásce (tj. po hlásce *k*). Konkrétně nás zajímá, zda se zvýrazní kontrast mezi

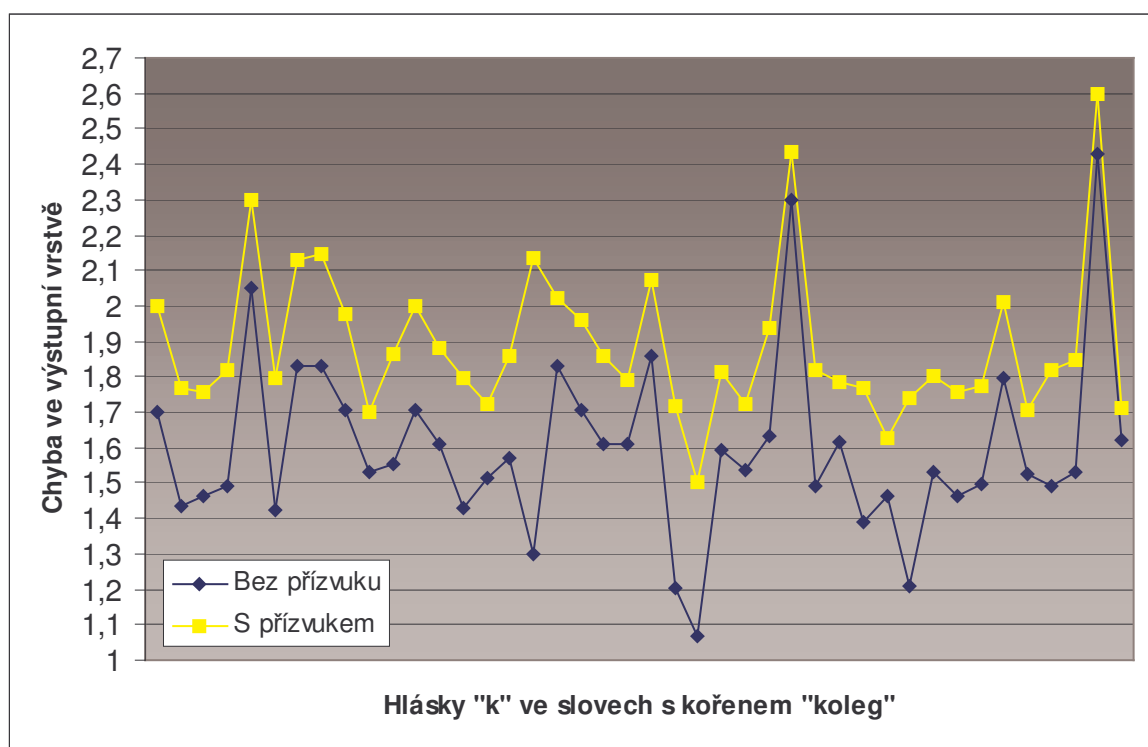
podmíněnou pravděpodobností výskytu první hlásky a průměrnou podmíněnou pravděpodobností výskytu hlásek, které tuto první hlásku bezprostředně následují, poté, co do tréninkového prostředí UNS přidáme informaci o umístění stálého přízvuku na první slabice slov s kořenem *koleg*.

Již z letmého pohledu na graf 36, ve kterém jsou zachyceny velikosti těchto rozdílů ve dvou různých experimentálních podmínkách, je dobře patrné, že s přidáním prozodického vodítka do tréninkového korpusu UNS se tento rozdíl u všech 42 slov s kořenem *koleg* zvětšuje. Jednostranný test statistické signifikance nárůstu tohoto rozdílu mezi chybovostí UNS při predikci první hlásky a průměrnou chybovostí UNS při predikci čtyř bezprostředně následujících hlásek prokázal, že na hladině významnosti $p < 0,0001$ můžeme odpovídající nulovou hypotézu zamítnout a konstatovat, že přidání prozodického vodítka do tréninkového prostředí UNS statisticky signifikantně zvětšuje („vyostřuje“) kontrast, který definuje hranici slova ve spojitém toku řeči, a tak i usnadňuje její identifikaci.



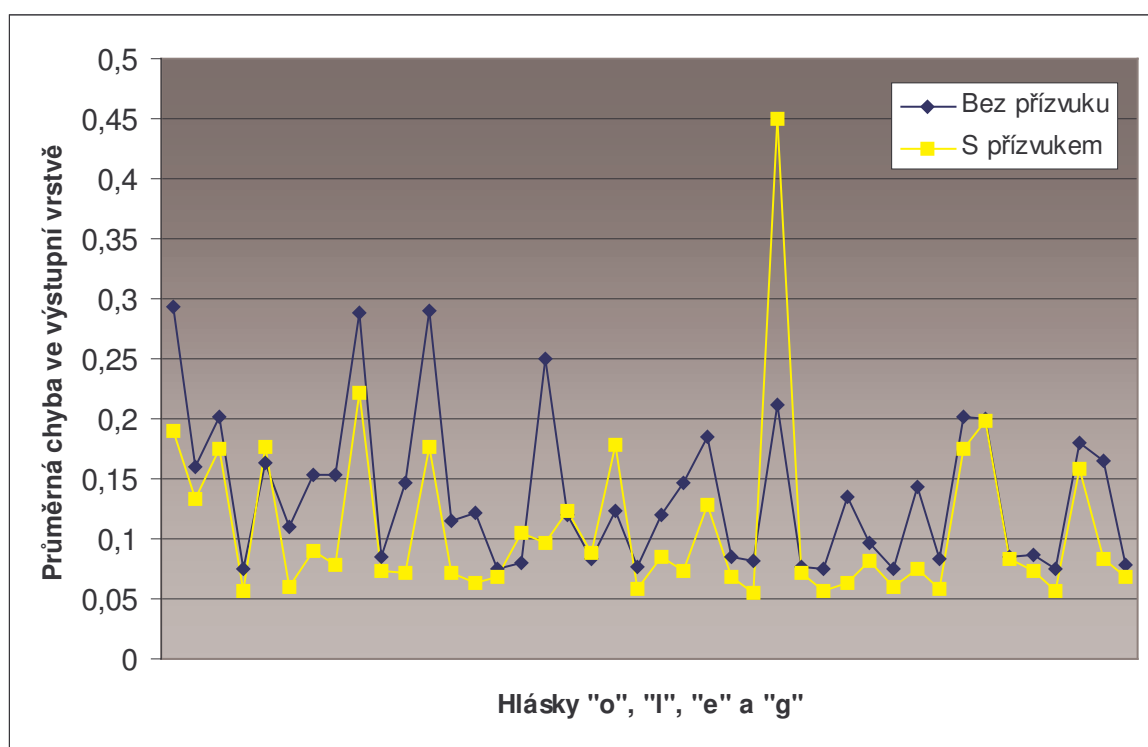
Graf 36: Velikost rozdílu mezi chybou UNS při predikci první hlásky a průměrnou chybou UNS při predikci čtyř bezprostředně následujících hlásek pro všech 42 slov s kořenem „koleg“ ve dvou různých experimentálních podmínkách, z nichž v jedné UNS má k dispozici informaci o umístění přízvuku na první slabice slov s kořenem „koleg“ a ve druhé ji k dispozici nemá.

Nyní, když víme, že přidání prozodického vodítka do tréninkového prostředí UNS statisticky signifikantně zvětšuje rozdíl mezi chybou při predikci první hlásky a průměrnou chybou při predikci čtyř bezprostředně následujících hlásek, můžeme přistoupit k testování těch hypotéz, jejichž předmětem je konkrétní mechanismus, který za tímto zvětšením rozdílu stojí. Jedním z možných mechanismů je zvýšení chybovosti UNS při predikci první hlásky slov s kořenem *koleg* (tj. při anticipaci hlásky *k*). Podle nulové hypotézy není z hlediska chyby, kterou UNS vykazuje při predikci hlásky *k*, statisticky významný rozdíl mezi situací, kdy informace o umístění přízvuku je součástí tréninkového prostředí, a situací, kdy tato informace součástí tréninkového prostředí není (viz Hypotéza 2). Stejně jako v předchozím případě, i zde můžeme nulovou hypotézu zamítnout ($p < 0,0001$) a konstatovat, že vložení informace o umístění přízvuku na první slabice slov s kořenem *koleg* do tréninkového prostředí UNS statisticky signifikantně zvýšilo chybovost („nejistotu“) UNS při anticipaci první hlásky všech slov s kořenem *koleg*, a že se tudíž tento mechanismus zcela jistě (spolu)podílí na „vyostření“ kontrastu mezi chybovými signály spojenými s predikcí první hlásky a čtyř bezprostředně následujících hlásek u všech slov s kořenem *koleg*.

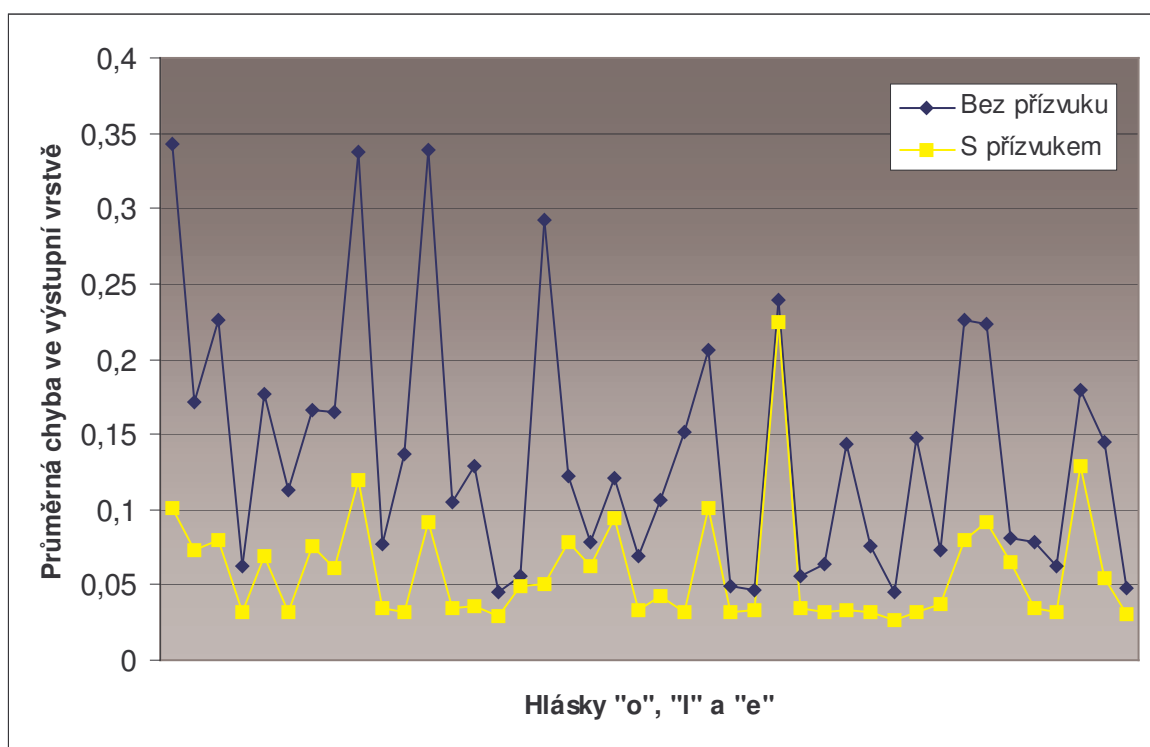


Graf 37: Chyba ve výstupní vrstvě pro všechny hlásky „k“, které stojí na začátku slov s kořenem „koleg“, ve dvou různých experimentálních podmínkách, z nichž v jedné UNS má k dispozici informaci o umístění přízvuku na první slabice slov s kořenem „koleg“ a ve druhé ji k dispozici nemá.

Druhým uvažovaným mechanismem bylo snížení průměrné chybovosti UNS při predikci čtyř hlásek, které bezprostředně následují po hlásce *k*. Existence tohoto konkrétního mechanismu je předmětem Hypotézy 3. Z grafu 38 je dobře patrné, že v naprosté většině případů je chybovost UNS při predikci čtyř hlásek následujících po hlásce *k* nižší v situaci, kdy je součástí tréninkového prostředí UNS informace o přízvuku. Tento pokles v chybovosti UNS by byl ještě výraznější, kdybychom průměrnou chybovost počítali pouze z prvních tří hlásek bezprostředně následujících po hlásce *k*, a nikoli ze všech čtyř kořenových hlásek (viz graf 39). Jednostranný test statistické signifikance potom jen potvrdil, že na hladině významnosti $p < 0,0001$ má vložení prozodického vodítka (v podobě přízvuku na první slabice slov s kořenem *koleg*) do tréninkového prostředí UNS za následek snížení její průměrné chybovosti při anticipaci čtyř hlásek následujících po hlásce *k* ve slovech s kořenem *koleg*. To potom tedy znamená, že na „vyostření“ kontrastu mezi chybovými signály spojenými s predikcí první hlásky a čtyř bezprostředně následujících hlásek u všech slov s kořenem *koleg* se spolupodílí i tento mechanismus.



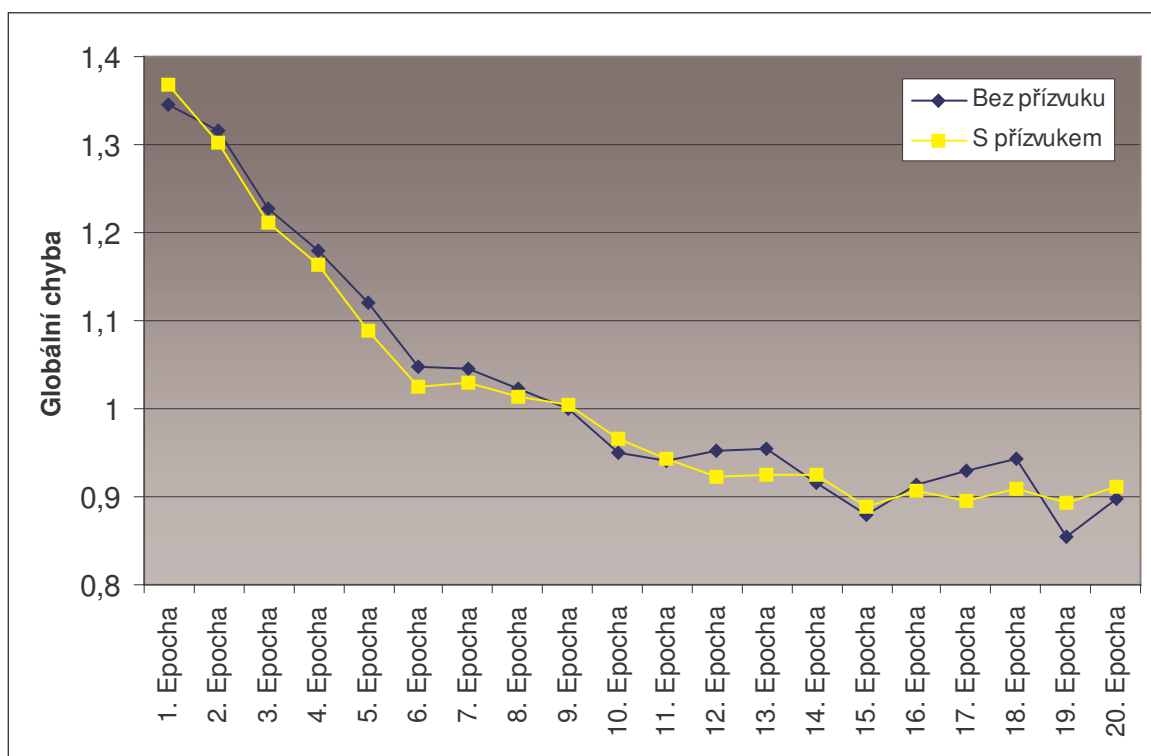
Graf 38: Průměrná chyba UNS při anticipaci čtyř hlásek, které následují po první hlásce ve slovech s kořenem „koleg“, ve dvou různých experimentálních podmínkách, z nichž v jedné má UNS k dispozici informaci o umístění přízvuku na první slabice slov s kořenem „koleg“ a ve druhé ji k dispozici nemá.



Graf 39: Průměrná chyba UNS při anticipaci tří hlásek, které následují po první hláске ve slovech s kořenem „koleg“, ve dvou různých experimentálních podmínkách, z nichž v jedné má UNS k dispozici informaci o umístění přízvuku na první slabice slov s kořenem „koleg“ a ve druhé nikoli.

Poslední, čtvrtá hypotéza se týká možnosti, že efekt dvou výše zmíněných (a potvrzených) mechanismů vzniku chybového kontrastu signalizujícího hranici slov se navzájem vyruší, takže na úrovni globální chyby, která je celkovým ukazatelem chybovosti UNS napříč celým tréninkovým korpusem, se výkon UNS exponované dvěma různým tréninkovým prostředím (s prozodickým vodítkem a bez něj) nebude statisticky významně lišit. Tuto domněnku se zdají podporovat grafy 26 a 27, na kterých je zachycen vývoj křivky globální chyby UNS během 20 tréninkových epoch ve dvou různých experimentálních podmínkách. Na první pohled se obě křivky od sebe téměř vůbec neliší. To však může být způsobeno pouze příliš hrubým měřítkem, ve kterém jsou naměřené hodnoty zobrazeny. Ke zjištění vlivu prozodického vodítka na celkovou chybovost UNS napříč celým tréninkovým korpusem byly použity údaje o globální chybě na konci každé tréninkové epochy, kdy UNS prošla celou množinou tréninkových dat. Pro obě UNS, resp. pro obě experimentální situace (s přízvukem a bez přízvuku) zde tak tedy bylo k dispozici 20 různých údajů, které bylo možné mezi sebou porovnávat. Tyto údaje jsou zaneseny v grafu 40. Friedmanovým testem bylo potvrzeno, že globální chyba UNS se v průběhu prvních 20 tréninkových epoch ve dvou různých experimentálních situacích statisticky

významně neliší ($Q = 0,8$; $df = 1$; $p = 0,371$), tzn. že přítomnost prozodického vodítka v tréninkovém prostředí UNS neměla (alespoň tedy v průběhu prvních 20 tréninkových epoch) za následek statisticky významné snížení nebo zvýšení celkové chybovosti UNS napříč celým tréninkovým korpusem.



Graf 40: Porovnání globální chyby na konci každé z 20 prvních tréninkových epoch ve dvou různých experimentálních situacích

Pro ověření robustnosti vlivu prozodického vodítka na činnost UNS byly provedeny ještě další dva nezávislé simulační experimenty, které se od původního simulačního experimentu lišily v několika málo slovech, která byla navíc součástí tréninkového prostředí.¹ U obou nových simulací se potvrdily všechny předchozí závěry založené na údajích z původního simulačního experimentu, a to vždy na srovnatelných hladinách statistické významnosti.

¹ Do původního tréninkového prostředí byla nejdříve přidána substantiva *pokora* a *lokomotiva* (ve všech svých sedmi pádech, v jednotném i množném čísle) a později další dvě substantiva *koza* a *škola* (opět ve všech svých sedmi pádech, v jednotném i množném čísle). V případě druhé jmenované dvojice slov byla součástí jejich reprezentačního vzorce rovněž informace o umístění přízvuku na jejich první slabice. Cílem tohoto rozšíření tréninkového prostředí UNS bylo zjistit, zda nedojde k nějaké významné změně ve vlivu prozodického vodítka na chování UNS při anticipaci hlásek slov s kořenem *koleg*, když se v jejím tréninkovém prostředí bude nacházet větší počet slov obsahujících posloupnost hlásek *ko* (ať už na začátku slova nebo kdekoli jinde ve slově) a majících ve svém aktivačním vzorci zakódovanou informaci o umístění přízvuku na své první slabice. Soubory s použitými posloupnostmi hlásek se nachází na příloženém CD v adresáři *Hlásky_Exp* pod názvy *Exp1b* a *Exp1c*.

Výsledky simulačního experimentu tedy potvrzují, že přidáním prozodického vodítka do tréninkového prostředí skutečně dochází k „vyostření“ kontrastu mezi chybovostí UNS při predikci první a několika následujících hlásek nového slova, a to jak prostřednictvím snížení podmíněné pravděpodobnosti výskytu první hlásky tak i prostřednictvím zvýšení podmíněné pravděpodobnosti výskytu několika bezprostředně následujících hlásek.

Nezodpovězenou otázkou však zůstává, co konkrétně způsobuje, že přítomnost prozodického vodítka v tréninkovém prostředí má na chování UNS právě ten vliv, jaký má. Jinak řečeno, proč vůbec dochází k tomu, že se s vložením prozodického vodítka do tréninkového prostředí UNS snižuje podmíněná pravděpodobnost výskytu první hlásky a naopak se zvyšuje podmíněná pravděpodobnost výskytu několika bezprostředně následujících hlásek?

Abychom mohli vysvětlit tento vliv přítomnosti prozodického vodítka v jazykovém vstupu na chování UNS, musíme si uvědomit, že to, co se UNS snaží predikovat, nejsou hlásky, ale 26-ti místné posloupnosti jedniček a nul, které kódují zvukové charakteristiky jednotlivých hlásek, včetně ne/přítomnosti přízvuku. UNS se tedy v každém časovém okamžiku snaží na základě své předešlé zkušenosti s prezentovaným jazykovým prostředím anticipovat to, která z pozic 26-ti místného vektoru (reprezentujícího očekávaný příští vstup) bude mít hodnotu 1 a která 0. Jestliže se pak v tréninkovém prostředí UNS nenachází žádná informace o umístění přízvuku, potom to z hlediska predikce následujícího vstupu znamená, že ať už bude příští hláska jakákoli, UNS si může být vždy stoprocentně jistá tím, že na začátku žádného z aktivačních vektorů reprezentujících příští vstup nebude 1, resp. že první pozice v reprezentačním vektoru, která kóduje ne/přítomnost přízvuku, je nastavena u všech hlásek vždy na nulu (= nepřítomnost přízvuku):

0 ?

S vložením prozodického vodítka do tréninkového prostředí však UNS tuto svou jistotu ztrácí. Je to jako vědět, že v nějakém sáčku se nachází v rovnoměrném poměru kuličky čtyř různých barev, které jsou přitom všechny vyrobeny ze stejného materiálu, například ze skla. Je-li tomu skutečně tak, potom si člověk při každém zanoření ruky do sáčku může být jist tím, že každá kulička, kterou si takto vytáhne, bude skleněnka, resp. že má pětadvacetiprocentní pravděpodobnost, že se strefí a uhodne, jaký konkrétní typ kuličky (daný kombinací barvy a materiálu, ze kterého je vyrobena) si vytáhne. Když se ale složení kuliček v sáčku změní tak, že jedna polovina z nich bude ze skla a druhá polovina z nějakého jiného materiálu, například

z kovu, potom to do procesu hádání vnáší nový prvek nejistoty, který se projeví snížením pravděpodobnosti, že člověk správně dopředu uhodne konkrétní typ vytáhnuté kuličky, na 12,5 %. Stejný princip můžeme aplikovat rovněž na chování UNS: Skutečnost, že jen některé hlásky ne/jsou spojené s přízvukem (tedy že jen u některých hlásek je odpovídající místo v jejich reprezentačním vektoru nastaveno na nulu) vnáší do anticipační úlohy UNS více nejistoty, která se obráží v její zvýšené chybovosti (nebo jinak formulováno, ve snížené pravděpodobnosti úspěchu UNS) při predikci dalšího vstupu.¹ Tímto konkrétním způsobem lze vysvětlit zvýšení chybovosti UNS při anticipaci první hlásky nového slova. Na základě obdobného principu však můžeme vysvětlit také její snížení při predikci hlásky, která bezprostředně následuje po první hlásce.

V tomto druhém případě dodání prozodického vodítka do tréninkového prostředí UNS nevede ke zvýšení, ale naopak ke snížení míry nejistoty ohledně toho, jaká další hláska bude následovat. Tento efekt je způsoben tím, že z hlediska UNS má přidání přízvuku do jazykového prostředí za následek vznik zcela nových hlásek, takže UNS se pak v průběhu tréninku učí mezi sebou asociovat množinu hlásek, která je poněkud odlišná od té, které je UNS vystavena v situaci, kdy informace o umístění přízvuku není součástí tréninkového prostředí. V případě, že za nově vzniklou hláskou s přízvukem se v daném jazykovém prostředí nachází menší počet různých hlásek než za stejnou hláskou, ale bez přízvuku, potom přidání prozodického vodítka logicky vede ke snížení nejistoty ohledně identity hlásky, která bezprostředně následuje po první hlásce nového slova. Takto například v rámci námi provedeného simulačního experimentu vznikla nová hláska *k* s přízvukem (dále označovaná jako *k̂*). Přitom po této konkrétní hlásce v tréninkovém korpusu vždy následovala hláska *o*, resp. *o* s přízvukem (*ô*). Naproti tomu v situaci, kdy součástí tréninkového prostředí nebylo prozodické vodítko, po hlásce *k* (bez přízvuku a na všech možných pozicích v různých slovech) mohlo následovat hned pět různých hlásek (*a*, *á*, *é*, *í* a *o*). Z toho je zřejmé, že když UNS byla na vstupu prezentována hláska *k̂*, potom si mohla být na základě své vlastní jazykové zkušenosti jistá, že po této konkrétní hlásce bude vždy následovat hláska *ô*. Naproti tomu v případě stejného slova a stejné hlásky, ovšem již bez přízvuku, se při predikci následující hlásky nutně musela projevit zkušenost UNS s jazykovým prostředím, kde po

¹ Tato zvýšená chybovost UNS je přitom přímým odrazem její tréninkové zkušenosti s jazykovým prostředím, které má určitou specifickou statistickou strukturu (ovlivněnou přítomností informace o umístění stálého přízvuku na první slabice), která se díky schopnosti UNS se učit zrcadlí v její váhové matici, a v důsledku toho tedy také v jejím chování (spočívajícím v predikci dalšího vstupu).

hlásce *k* (bez přízvuku) následovalo pět různých hlásek, a to v (mírně)¹ zvýšené chybovosti („nejistotě“) UNS při predikci jejího příštího vstupu

Z výše uvedené úvahy také vyplývá, že vliv prozodického vodítka na míru chybovosti UNS při anticipaci hlásky bezprostředně následující po první hlásce je kriticky závislý na skutečnosti, že za nově vzniklou hláskou s přízvukem se v daném jazykovém prostředí nachází menší počet různých hlásek než za stejnou hláskou, ale bez přízvuku, což je předpoklad, který je v případě jenom trochu reprezentativnějšího vzorku českého jazyka (kde by se na pozici těsně za první hláskou s přízvukem prostřídalo větší množství různých hlásek) dost nereálný. Proto lze vyslovit hypotézu, že v realističtějších podmínkách by vliv přítomnosti prozodického vodítka v tréninkovém prostředí na chování UNS – v podobě vyostření kontrastu mezi její chybovostí při predikci první a bezprostředně následující hlásky – byl zprostředkován především (nebo výlučně) zvýšenou chybovostí UNS při predikci první hlásky.

K ověření této hypotézy byl proveden samostatný simulační experiment, který se od toho původního lišil pouze ve dvou klíčových parametrech: V počtu epoch během tréninkové fáze experimentu, kterých nebylo 20, ale 40, a potom ve složení tréninkového korpusu. V tom bylo ve zcela náhodném pořadí obsaženo 14 různých substantiv, vždy v prvním pádě jednotného čísla (kapitán, herpes, prase, chalupa, cihla, závit, lano, průkaz, násyp, marnost, dohled, jasan, vidle a fáma); od každého slova bylo v korpusu 7 exemplářů. V tréninkovém prostředí se tedy celkem nacházelo 98 slov, která po rozložení na lineární sled hlásek vytvořila ničím nepřerušovanou posloupnost 532 hlásek.² Stejně jako v prvním simulačním experimentu i zde byla UNS v průběhu svého učení vystavena dvěma různým tréninkovým prostředím, z nichž pouze v jednom byla obsažena informace o umístění stálého přízvuku na první slabice slova *kapitán*. Z hlediska testování výše uvedené hypotézy byla klíčovou vlastností tréninkového korpusu přítomnost pouze dvou slov, která v sobě obsahovala hlásku *k* (v případě *kapitána* na začátku slova a v případě *průkazu* uprostřed), po které vždy následovala hláska *a*. V důsledku toho by UNS neměla mít nijak usnadněnou úlohu predikce hlásky *a* nacházející se za hláskou *k* ve slově *kapitán* poté, co by do jejího tréninkového prostředí bylo vloženo prozodické vodítko. Jinak řečeno, UNS by v této experimentální situaci neměla vykazovat sníženou chybovost při anticipaci této hlásky, ba naopak, tato chybovost by se měla mírně zvýšit, neboť

¹ Zvýšení chybovosti UNS není sice nijak dramatické (neboť UNS díky své kontextové vrstvě a specifickému způsobu zpracovávání informací při anticipaci následujícího vstupu nezohledňuje pouze výskyt bezprostředně předcházející hlásky, ale de facto všech předcházejících hlásek), ale ze získaných dat přesto jasně patrné (viz obrázky 36 a 37).

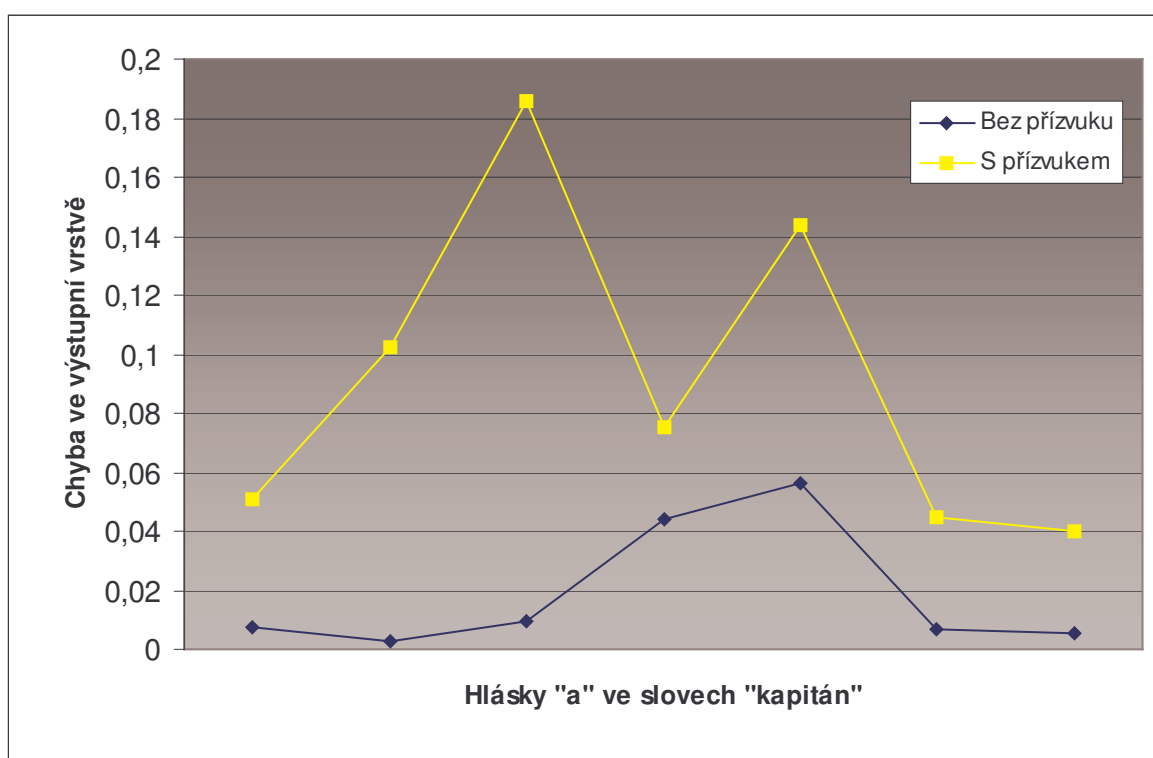
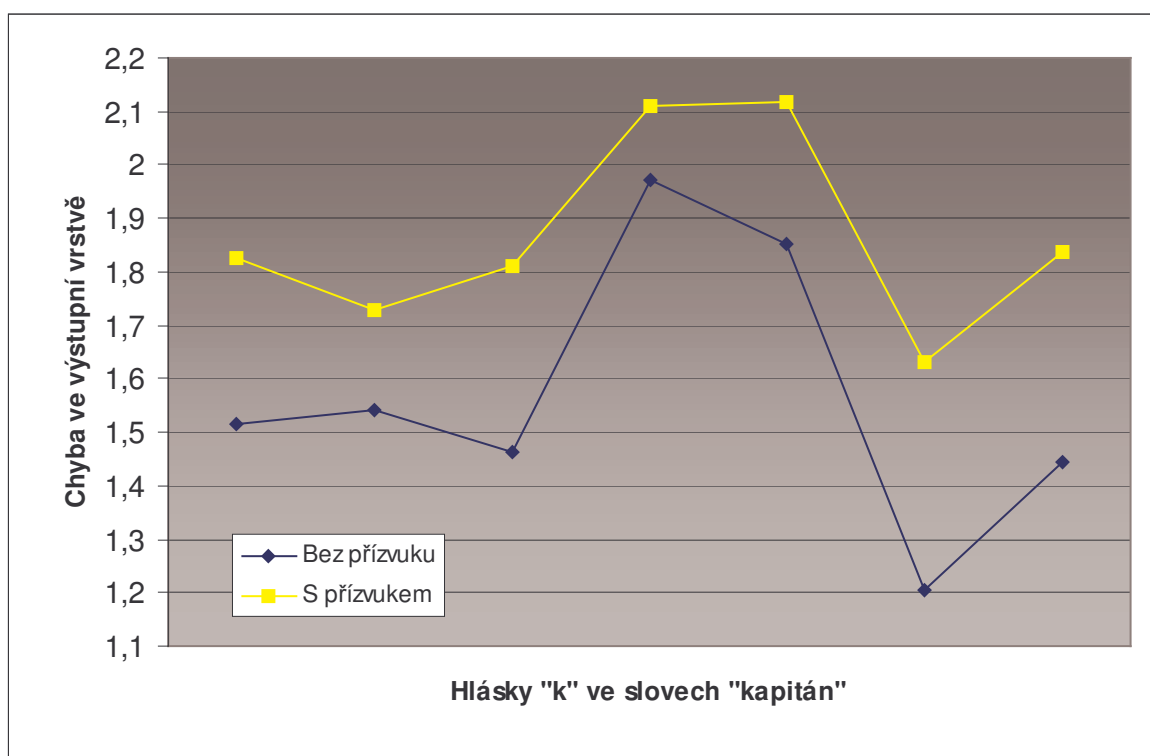
² Soubor s posloupností těchto hlásek se nachází na příloženém CD v adresáři *Hlásky_Exp* pod názvem *Exp2*.

vložení prozodického vodítka do tréninkového prostředí vzniknou de facto dvě zcela nové hlásky – *k* a *a* s přízvukem (*ḳ*, *ạ*) -, čímž se sníží původní počet 14 (2 x 7) dvojic hlásek *ka*, které má UNS šanci „potkat“ během jedné tréninkové epochy, na 7, a takové „setkání“ pro UNS vždy představuje příležitost něčemu novému se naučit, resp. příležitost k úpravě stávajících vážených spojů tak, aby se snížila chybovost ve výstupní vrstvě při prezentaci daného vstupu. Ze stejného důvodu by po vložení přízvuku do tréninkového prostředí mělo dojít také k mírnému zvýšení chybovosti UNS při predikci hlásky *a* nacházející se za hláskou *k* ve slově *průkaz*. Z toho všeho pak vyplývá, že k tomu, aby se zachoval přínos prozodického vodítka k segmentaci řeči¹, musí být nárůst v chybovosti UNS při predikci první hlásky větší než nárůst chybovosti UNS při predikci hlásky, která ji bezprostředně následuje. Na základě výše uvedených úvah můžeme tedy formulovat pět následujících hypotéz:

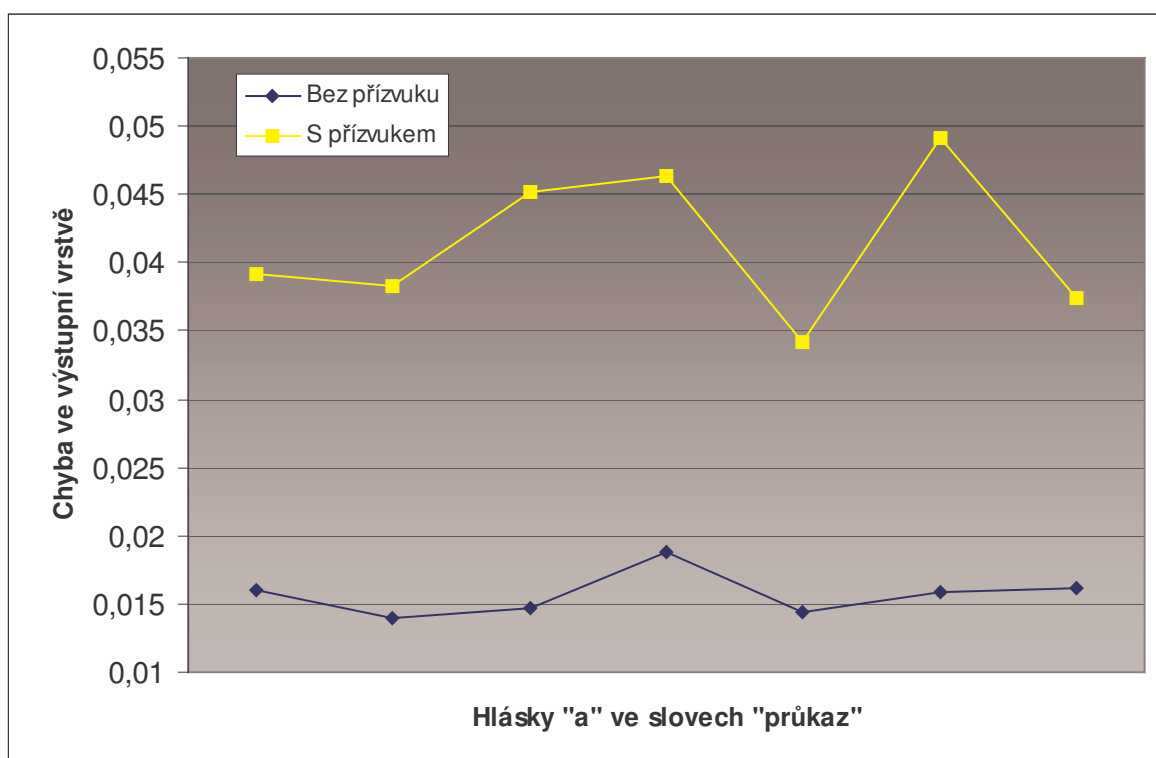
- 1) S přidáním prozodického vodítka do tréninkového prostředí UNS stoupne její chybovost při predikci hlásky *k* ve slovech *kapitán*.
- 2) S přidáním prozodického vodítka do tréninkového prostředí UNS stoupne její chybovost při predikci hlásek *a* ve slovech *kapitán*.
- 3) S přidáním prozodického vodítka do tréninkového prostředí UNS stoupne její chybovost při predikci hlásek *a* ve slovech *průkaz*.
- 4) S přidáním prozodického vodítka do tréninkového prostředí UNS stoupne její chybovost při predikci hlásek *k* více než při predikci hlásek *a* ve stejných slovech.
- 5) S přidáním prozodického vodítka do tréninkového prostředí UNS se zvětší rozdíl mezi chybovostmi, kterou UNS vykazuje při predikci hlásek *k* a *a* ve slovech *kapitán*.

Po provedení samotné simulace a shromáždění všech potřebných údajů, které jsou zaneseny v grafech 40-44, bylo všech pět hypotéz testováno pomocí Wilcoxonova neparametrického pořadového testu pro párované hodnoty, vždy s jednostrannou signifikancí a s požadovanou hladinou významnosti $\alpha = 0,01$. Zjištěné údaje potvrdily všech pět výše uvedených hypotéz na hladině významnosti $p < 0,008$. Tyto výsledky tak naznačují, že v situaci, kdy by UNS byla během procesu učení vystavena reprezentativnějšímu vzorku českého jazyka, ve kterém by se na pozici těsně za první hláskou s přízvukem prostřídalo stejné (nebo alespoň srovnatelné) množství různých hlásek jako za stejnou hláskou bez přízvuku, potom by přínos prozodického vodítka (v podobě stálého přízvuku na první slabice slova) pro segmentaci spojitého toku řeči byl skutečně zprostředkován především (nebo výhradně) nárůstem chybovosti UNS při anticipaci první hlásky nového slova.

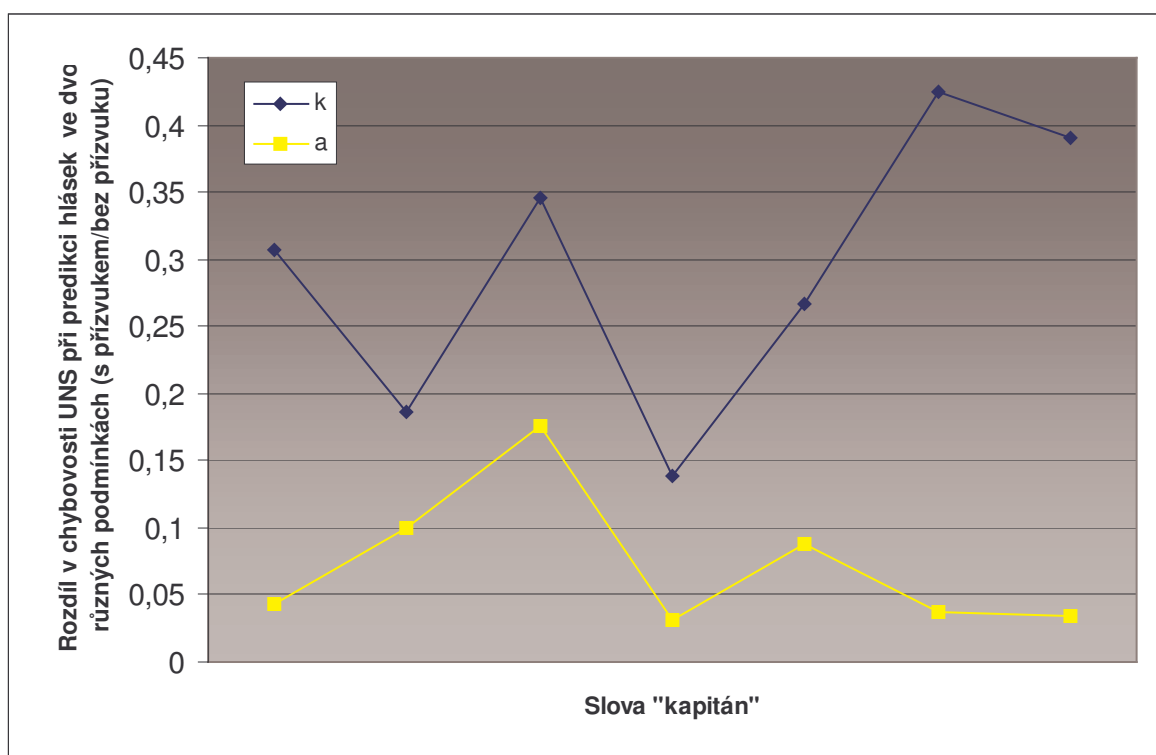
¹ Přínos v podobě vyostření kontrastu mezi chybovostmi UNS při predikci první a bezprostředně následující hlásky.



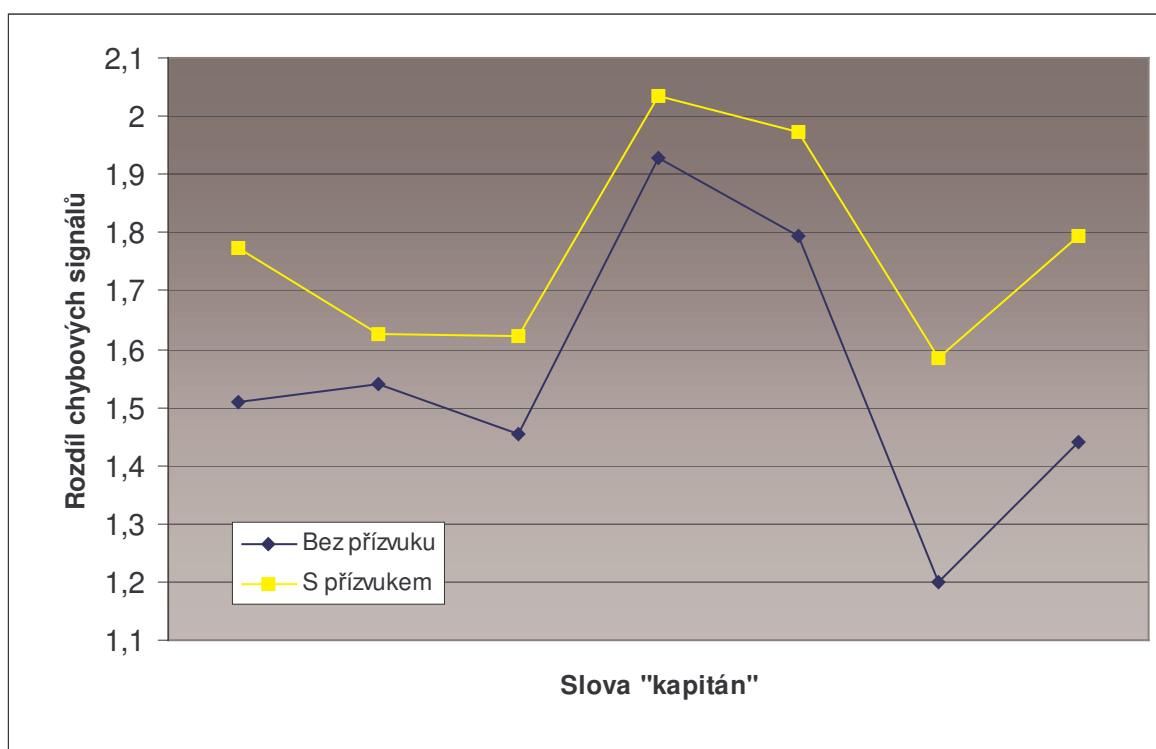
Graf 40-41: Chyba ve výstupní vrstvě pro všechny hlásky „k“ a „a“, které stojí na začátku slov „kapitán“, ve dvou různých experimentálních podmínkách, z nichž pouze v jedné má UNS k dispozici informaci o umístění přízvuku na první slabice slov „kapitán“.



Graf 42: Chyba ve výstupní vrstvě pro všechny hlásky „a“, které stojí na začátku slov „průkaz“, ve dvou různých experimentálních podmínkách (s přízvukem/bez přízvuku na první slabice slov „kapitán“).



Graf 43: Porovnání rozdílů mezi chybovostí UNS při predikci hlásek „k“ a „a“ ve dvou různých experimentálních situacích (s přízvukem/bez přízvuku na první slabice slov „kapitán“).



Graf 44: Velikost rozdílu mezi chybou UNS při predikci první hlásky („k“) a chybou UNS při predikci bezprostředně následující hlásky („a“) pro všech sedm slov „kapitán“ ve dvou různých experimentálních podmínkách (s přízvukem/bez přízvuku na první slabice slov „kapitán“).

4.7 Závěr

Výsledky prvního simulačního experimentu potvrzují Elmanovo původní zjištění, že jednoduchá rekurentní síť anticipující svůj příští vstup dokáže v jazykovém vstupu objevit jistý druh statistických zákonitostí v souvýskytu jednotlivých hlásek, které se projevují v její rozdílné chybovosti při predikci jednotlivých hlásek podle toho, na jaké pozici ve slově se tyto hlásky nacházejí. Pro první hlásku každého slova je takto charakteristická relativně vysoká míra chybovosti UNS při její predikci, která směrem od ní postupně klesá a poté opět prudce stoupá na začátku dalšího slova. Tímto způsobem jsou jasně vyznačeny hranice jednotlivých slov ve spojitém toku řeči, tvořeném nepřerušovanou posloupností hlásek, jako ta místa, kde je nízká podmíněná pravděpodobnost výskytu další hlásky za předpokladu výskytu všech předcházejících hlásek a kde se ztrácí jakákoli souvislost mezi danou hláskou a všemi předcházejícími hláskami.

Použití flektivního jazyka jakým je čeština jako vstupu do jednoduché rekurentní sítě prokázalo, že obdobný princip je možné využít rovněž při lokalizaci hranic mezi kořenem slova a jeho gramatickou koncovkou, čímž se tento princip může spolupodílet na procesu

osvojování jazykové flexe, která na straně kompetentního mluvčího předpokládá určitou schopnost odlišit od sebe obsahové a gramatické morfémy.

Hranice slova v Elmanově pojetí však neumožňuje ve spojitém toku řeči odlišit začátek slova od gramatické koncovky, neboť tyto dvě části slova se velmi často vyznačují podobně nízkou podmíněnou pravděpodobností výskytu. Tento problém řeší nově definovaná hranice slova chápáná nyní jako místo výrazného kontrastu (rozdílu) mezi podmíněnou pravděpodobností výskytu první hlásky a jedné nebo více bezprostředně následujících hlásek.

V kontextu takto nově chápané hranice slov výsledky prvního simulačního experimentu prokázaly, že přidání prozodického vodítka do tréninkového prostředí UNS statisticky signifikantně zvětšuje („vyostřuje“) kontrast, který definuje hranici slova ve spojitém toku řeči, a tak i usnadňuje její identifikaci.

Další analýzou vlivu prozodického vodítka na chování UNS se prokázalo, že výše zmíněný efekt byl způsobem dvěma různými mechanismy: Zvýšením chybovosti UNS při predikci první hlásky a naopak snížením chybovosti UNS při anticipaci jedné nebo více hlásek, které po první hlásce bezprostředně následují. Zatímco v prvním případě přidání prozodického vodítka do tréninkového prostředí vneslo do anticipační úlohy nový prvek nejistoty, v druhém případě tomu bylo přesně naopak.

Druhý simulační experiment ukázal, že v případě druhého mechanismu by se mohlo jednat pouze o artefakt nereprezentativního složení tréninkového prostředí a že při použití reprezentativnějšího vzorku českého jazyka, ve kterém by se na pozici těsně za první hláskou s přízvukem prostřídalo stejné (nebo alespoň srovnatelné) množství různých hlásek jako za stejnou hláskou bez přízvuku, by byl přínos prozodického vodítka pro segmentaci spojitého toku řeči zprostředkován především (nebo výhradně) nárůstem chybovosti UNS při anticipaci první hlásky nového slova.

Na úplný závěr můžeme konstatovat, že výsledky provedených simulací naznačují, že Elmanova jednoduchá rekurentní síť ve spojení s nově definovanou hranicí slov představuje celkem plausibilní model segmentace řeči, který dokáže v rámci jednoho konceptuálního rámce implementovat mechanismus exploatace fonotaktického a prozodického vodítka, o kterých se předpokládá, že je malé děti využívají při segmentaci spojitého toku řeči na jednotlivá slova. Toto tvrzení však nelze chápat tak, že by mozek malého dítěte při segmentaci řeči pracoval naprosto stejným způsobem jako jednoduchá rekurentní síť. Jedná se pouze o konstatování, že jistý typ paralelně distribuovaného výpočetního zařízení, jehož konektivita do určité míry připomíná tu, kterou můžeme nalézt v kůře lidského mozku, dokáže čistě na základě své zkušenosti s jazykovým prostředím objevit v tomto prostředí jisté

statistické zákonitosti, které je možné využít při segmentaci řeči. Použijeme-li v této souvislosti již několikrát zmíněnou horolezeckou metaforiku, potom to znamená, že jednoduchá rekurentní síť (jakožto velice zjednodušený model lidského mozku) disponuje dostatečně kvalitní horolezeckou výstrojí, která jí umožňuje zdolat zdánlivě zcela hladkou stěnu spojitého toku řeči, a že prozodické vodítko v podobě stálého přízvuku na první slabice slova má za následek zvětšení ve stěně se přirozeně vyskytujících trhlin, kterých se tak lze při slézání stěny mnohem snáz zachytit.

Seznam literatury

- Adams, D. (2002): *Život, vesmír a vůbec*. Argo, Praha
- Abraham, F. D. (1995): *Introduction to Dynamics: A Basic Language; A Basic Metamodeling Strategy*. In Abraham, F. D.; Gilgen, A. R. (Eds.): *Chaos Theory In Psychology*. Praeger, London
- Altmann, G. T. M. (2005): *Výstup na babylonskou věž. Otázky jazyka, mysli a porozumění*. Triáda, Praha
- Altmann, G. T. M. (2002): *Statistical learning in infants*. In *PNAS*, 99 (24), s. 15250-15251
- Altmann, G. T. M. (2001): *The language machine: Psycholinguistics in review*. In *British Journal of Psychology*, 92, s. 129-170
- Axelrod, R. (1976): *Structure of Decision: The Cognitive Maps of Political Elites*. Princeton University Press, New Jersey
- Baillargeon, R. (1993): *The object concept revisited. New directions in the investigation of infants' physical knowledge*. In Granrud, C. E. (Ed.): *Visual Perception and Cognition*. Erlbaum, Hillsdale, NJ
- Barabási, A.-L. (2005): *V pavučině sítí*. Paseka, Praha
- Beer, R. D. (2000): *Dynamical approaches to cognitive science*. In *Trends in Cognitive Sciences*, 4(3), s. 91-99.
- Bickerton, D. (1981): *Roots of language*. Karoma, Ann Arbor
- Blecha, I. (1996): *Filosofie. Základní problémy*. Nakladatelství FIN, Olomouc
- Borges, B.; Goldstein, D. G.; Ortmann, A.; Gigerenzer, G. (1999): *Can Ignorance Beat the Stock Market?* In Gigerenzer, G.; Todd, P. M. a kol.: *Simple Heuristics That Make Us Smart*. Oxford University Press, New York
- Braitenberg, V. (1984): *Vehicles: Experiments in Synthetic Psychology*. MIT Press, Cambridge, MA
- Brenkus, P. (2005): *Londýnská věž – neuropsychologický test plánování: Srovnání dvou verzí*. Diplomová práce, nevydáno, Katedra psychologie FF UK, Praha
- Broks, P. (2005): *Do země ticha. Lidský mozek – tajemný svět „uvnitř hlavy“*. NLN, Praha
- Brooks, R. A. (2003): *Robot: The Future of Flesh and Machines*. Penguin Group, New York
- Brooks, R. A. (1999): *Cambrian intelligence: the early history of the new AI*. MIT Press, Cambridge, MA

- Bruner, J. S. (1957): *Going beyond the information given*. In Gruber, H. E.; Hammond, K. R.; Jessor, R. (Eds.): *Contemporary approaches to cognition*. Harvard University Press, Cambridge, MA
- Budil, I. T. (1998): *Mýtus, jazyk a kulturní antropologie*. TRITON, Praha
- Burgess, C.; Lund, K. (1997): *Modelling parsing constraints with high-dimensional context space*. In *Language and Cognitive Processes*, 12, s. 177–210
- Buzan, T. (2002): *Síla kreativní inteligence*. Columbus, Praha
- Calvin, W. H. (2000): *Ako myslí mozog*. Kalligram, Bratislava
- Clark, A. (2003): *Natural-Born Cyborgs. Minds, Technologies, and the Future of Human Intelligence*. Oxford University Press, New York
- Clark, A. (2001): *Mindware. An Introduction to the Philosophy of Cognitive Science*. Oxford University Press, New York
- Clark, A. (1998a): *Magic Words: How language Augments Computation*. In Carruthers, P.; Boucher, J. (Eds.): *Language and Thought: Interdisciplinary Themes*. Cambridge University Press, Cambridge, s. 162-183
- Clark, A. (1998b): *Economic Reason: The Interplay of Individual Learning and External Structure*. In Drobak, J.; Nye, J. (Eds.): *The Frontiers Of The New Institutional Economics*. Academic Press, San Diego, CA, s. 269-290
- Clark, A.; Chalmers, D. J. (1998): *The Extended Mind*. In *Analysis*, 58 (1), s. 7-19
- Clark, A. (1997): *Being There. Putting Brain, Body, and World Together Again*. MIT Press, Cambridge, MA
- Clark, A. (1989): *Microcognition: Philosophy, Cognitive Science, and Parallel Distributed Processing*. MIT Press, Cambridge, MA
- Clayton, K. (1997): *Basic Concepts in Nonlinear Dynamics and Chaos*. [online] [cit. 1-3-2007] Dostupné z WWW: <<http://www.societyforchaostheory.org/chaosprimer.pdf>>
- Cooper, R. P.; Aslin, R. N. (1990): *Preference for infant directed speech in the first month after birth*. In *Child Development*, 61, s. 1584-1595
- Crutchfield, J. P.; Mitchell, M.; Das, R. (2003): *The evolutionary design of collective computation in cellular automata*. In Crutchfield, J. P.; Schuster, P. K. (Eds.): *Evolutionary Dynamics - Exploring the Interplay of Selection, Neutrality, Accident, and Function*. Oxford University Press, New York, s. 361-411
- Cumminsová, D. D. (1998): *Záhady experimentální psychologie. Co psychologové zjistili o myšlení, citech a chování člověka*. Portál, Praha
- Černý, J. (1998): *Úvod do studia jazyka*. Rubico, Olomouc

- Damasio, A. R. (2000): *Descartesův omyl. Emoce, rozum a lidský mozek*. Mladá fronta, Praha
- D'Andrade, R. (1995): *The Development of Cognitive Anthropology*. Cambridge University Press, Cambridge, MA
- Das, R.; Crutchfield, J. P.; Mitchell, M.; Hanson, J. E. (1995): *Evolving globally synchronized cellular automata*. In Eshelman, L. J. (Ed.): *Proceedings of the Sixth International Conference on Genetic Algorithms*. Morgan Kaufmann, San Mateo, CA
- Dawkins, R. (2002): *Slepý hodinář*. Paseka, Praha
- Dawkins, R. (1998): *Sobecký gen*. Mladá fronta, Praha
- Dawkins, R. (1996): *Řeka z ráje. Darwinistický pohled na život*. Archa, Bratislava
- Dawson, M. R. W. (2004): *Connectionism And Psychological Modeling*. Blackwell Publishers, Malden, MA
- Dawson, M. R. W. (2002): *From Embodied Cognitive Science To Synthetic Psychology*. In Wang, Y.; Johnston, R. H.; Smith, M. R. (Eds.): *Proceedings Of The First IEEE International Conference On Cognitive Informatics (ICCI'02)*. IEEE Computer Society, Los Alamitos, CA, s. 13-22
- DeCasper, A. J.; Spence, M. J. (1986): *Prenatal maternal speech influences newborns' perception of speech sounds*. In *Infant Behaviour and Development*, 9, s. 133–150
- Dennett, D. C. (2004a): *Freedom Evolves*. Penguin Group, New York
- Dennett, D. C. (2004b): *Druhy myslí. K pochopení vědomí*. Academia, Praha
- Dennett, D. C. (1998): *Brainchildren: Essays on Designing Minds*. MIT Press, Cambridge, MA
- Dennett, D. C. (1996): *Jak dělat chyby*. In Brockman, J.; Matsonová, K. (Eds.): *Jak se věci mají. Průvodce myšlenkami moderní vědy*. Archa, Bratislava
- Dennett, D. C. (1995): *Darwin's Dangerous Idea. Evolution and the Meanings of Life*. Simon & Schuster Paperbacks, New York
- Dennett, D. C. (1993): *Consciousness Explained*. Penguin Group, New York
- Dennett, D. C. (1987): *The Intentional Stance*. MIT Press, Cambridge, MA
- Dickerson, J. A.; Kosko, B. (1996): *Virtual Worlds as Fuzzy Dynamical Systems*. In Sheu, B. (Ed.): *Technology for Multimedia*. IEEE Press
- Donald, M. (1991): *Origins of the Modern Mind. Three Stages in the Evolution of Culture and Cognition*. Harvard University Press, Cambridge, MA
- Eco, U. (2000): *Svatá válka: MacIntosh v. DOS*. In *MF DNES*, 4.10., s. 20
- Edelman, G. M. (1987): *Neural Darwinism: The Theory of Neuronal Group Selection*. Basic Books, New York

- Eimas, P. D.; Siqueland, E. R.; Jusczyk, P.; Vigorito, J. (1971): *Speech perception in infants*. In *Science*, 171, s. 303–306
- Elman, J. L. (2004): *An alternative view of the mental lexicon*. In *Trends in Cognitive Sciences*, 7, s. 301-306
- Elman, J. L. a kol. (1996): *Rethinking Innateness: A Connectionist Perspective on Development*. MIT Press, Cambridge, MA
- Elman, J. L. (1996): *Learning rediscovered*. In *Science*, 274, s. 1849-1850
- Elman, J. L. (1995): *Language as a dynamical system*. In van Gelder, T.; Port, R. (Eds.): *Mind as Motion: Explorations in the Dynamics of Cognition*. MIT Press, Cambridge
- Elman, J. L. (1993): *Learning and development in neural networks: The importance of starting small*. In *Cognition*, 48, s. 71-99
- Elman, J. L. (1990): *Finding structure in time*. In *Cognitive Science*, 14, s. 179–211
- Fanta, J. (2001): *Psychologie, algoritmy a umělá inteligence na kapitálových trzích*. Grada, Praha
- Fanta, J. (2000): *Neuronové sítě ve společenských vědách*. Karolinum, Praha
- Fauconnier, G.; Turner, M. (2003): *The Way We Think. Conceptual Blending and the Mind's Hidden Complexities*. Basic Books, New York
- Felleman, D.; van Essen, D. (1991): *Distributed hierarchical processing in primate visual cortex*. In *Cerebral Cortex*, 1, 1-47
- Ferjenčík, J. (2000): *Úvod do metodologie psychologického výzkumu. Jak zkoumat lidskou duši*. Portál, Praha
- Fernald, A. (1992): *Human maternal vocalizations to infants as biologically relevant signals: an evolutionary perspective*. In Barkow, J. H.; Cosmides, L.; Tooby, J. (Eds.): *The adapted mind*. Oxford University Press, New York
- Fišar, Z.; Jiráček, R. (2001): *Vybrané kapitoly z biologické psychiatrie*. Grada Publishing, Praha
- Fodor, J. A.; Pylyshyn, Z. (1988): *Connectionism and cognitive architecture: A critical analysis*. In *Cognition*, 28, s. 3-71
- Fodor, J. A. (1987): *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. MIT Press, Cambridge, MA
- Fodor, J. A. (1983): *The modularity of mind*. MIT Press, Cambridge, MA
- Folta, J.; Loucká, P.; Klíma, B. (1997): *Věstonická vrubovka. Sloužil paleolitický předmět k bijekci mezi prvky dvou množin?* In *Vesmír*, 76, s. 310
- Franklin, S. (2001): *Artificial Minds*. MIT Press, Cambridge, MA

- van Gelder, T.; Port, R. (1995): *It's about time: An overview of the dynamical approach to cognition*. In van Gelder, T.; Port, R. (Eds.): *Mind as Motion: Explorations in the Dynamics of Cognition*. MIT Press, Cambridge
- Georgopoulos, V. C.; Malandraki, G. A.; Stylios, Ch. D. (2003): *A fuzzy cognitive map approach to differential diagnosis of specific language impairment*. In *Artificial Intelligence in Medicine*, 29, s. 261–278
- Gigerenzer, G. (2000): *Mind as Computer. The Social Origin of a Metaphor*. In Gigerenzer, G. (Ed.): *Adaptive Thinking. Rationality in the Real World*. Oxford University Press, New York
- Gigerenzer, G.; Todd, P. M. a kol. (1999): *Simple Heuristics That Make Us Smart*. Oxford University Press, New York
- Gilovich, T.; Griffin, D.; Kahneman, D. (Eds.) (2002): *Heuristics and Biases. The Psychology of Intuitive Judgment*. Oxford University Press, New York
- Gleick, J. (1996): *Chaos*. Ando Publishing, Brno
- Gleitman, L. (1990): *The structural sources of verb meanings*. In *Language Acquisition*, 1, s. 3–55
- Gold, E. M. (1967): *Language identification in the limit*. In *Information and Control*, 16, s. 447-474
- Gomez, R. L.; Gerken, L. A. (1999): *Artificial grammar learning by one-year-olds leads to specific and abstract knowledge*. *Cognition*, 70 (2), s. 109–136
- Greene, B. (2006): *Struktura vesmíru. Čas, prostor a povaha reality*. Paseka, Praha
- Greene, B. (2002): *Elegantní vesmír*. Mladá fronta, Praha
- Gribbin, J. (1998): *Pátrání po Schroedingerově kočce. Kvantová fyzika a skutečnost*. Columbus, Praha
- Haugeland, J. (1997): *What is mind design?* In Haugeland, J. (Ed.): *Mind Design II*. MIT Press, Cambridge, MA
- Haugeland, J. (1981): *Semantic engines. An introduction to mind design*. In Haugeland, J. (Ed.): *Mind Design*. MIT Press, Cambridge, MA
- Havel, I. M. (2000): *Věda o duši*. In *Vesmír*, 79, s. 363
- Hebb, D. (1949): *The Organization of Behavior*. Wiley, New York
- Hillis, W. D. (2003): *Vzor v kameni. Jednoduché myšlenky, které řídí počítače*. Academia, Praha
- Hoernes, G. E.; Heilweil, M. F. (1969): *Úvod do Booleovy algebry a navrhování logických obvodů*. SNTL, Praha

- Hofstadter, D. (2000): *Gödel, Escher, Bach: an Eternal Golden Braid*. Penguin Group, New York
- Hofstadter, D. (1985): *Metamagical Themas: Questing for the Essence of Mind and Pattern*. Penguin Group, New York
- Hoskovec, J.; Nakonečný, M.; Sedláková, M. (2002): *Psychologie XX. století*. Karolinum, Praha
- Huerga, A. V. (2005): *A Balanced Differential Learning Algorithm in Fuzzy Cognitive Maps*. [online] [cit. 1-3-2007] Dostupné z WWW: <http://monet.aber.ac.uk:8080/monet/docs/pdf_files/gr_02/gr2002albertovazques.pdf>
- Hume, D. (1972). *Zkoumání lidského rozumu*. Svoboda, Praha
- Husserl, E. (1996): *Přednášky k fenomenologii vnitřního časového vědomí*. Ježek, Praha
- Hutchins, E. (2004): *Material Anchors for Conceptual Blends*. [online] [cit. 31-10-2006] Dostupné z WWW: <http://hci.ucsd.edu/lab/hci_papers/EH2004-1.pdf>
- Hutchins, E. (1995): *Cognition in the Wild*. MIT Press, Cambridge, MA
- Chomsky, N. (1980): *Rules and Representations*. Columbia University Press, New York
- Chomsky, N. (1975): *Reflections on Language*. Pantheon Books, New York
- Chomsky, N. (1959): Review of Skinner's *Verbal behaviour*. In *Language*, 35, s. 26–58
- Chomsky, N. (1957): *Syntactic structures*. Mouton, The Hague
- Churchland, P. C.; Sejnowski, T. J. (1999): *The Computational Brain*. MIT Press, Cambridge, MA
- Churchland, P. M. (1981): *Eliminative Materialism and Propositional Attitudes*. In *The Journal of Philosophy*, 78 (2), s. 67-90
- Imai, M.; Gentner, D.; Uchida, N. (1994): *Children's theories of word meaning: The role of shape similarity in early acquisition*. In *Cognitive Development*, 9, s. 45–75
- Jusczyk, P. W. (1997): *The discovery of spoken language*. MIT Press, Cambridge, MA
- Jusczyk, P. W.; Aslin, R. N. (1995): *Infants' detection of the sound patterns of words in fluent speech*. In *Cognitive Psychology*, 29, s. 1-23
- Kahneman, D.; Slovic, P.; Tversky, A. (1982): *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press: Cambridge, MA
- Kelemen, J.; Kubík, A.; Lenharčík, I.; Mikulecký, P. (1999): *Tvorba expertních systémů v prostředí CLIPS*. Grada, Praha
- Kelemen, J. (1995): *Budoucí Altamira*. Votobia, Praha
- Kelemen, J. (1994): *Strojovia a agenty*. Archa, Bratislava
- Kim, J. (1996): *Philosophy of Mind*. Westview Press, Boulder, CO

- Kirsh, D. (1996): *Adapting the Environment Instead of Oneself*. In *Adaptive Behavior*, 4, s. 415-452
- Kirsh, D. (1995a): *Complementary Strategies: Why we use our hands when we think*. In Moore, J. D.; Lehman, J. F. (Eds.): *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society*. s. 212-217
- Kirsh, D. (1995b): *The Intelligent Use of Space*. In *Artificial Intelligence*, 73 (1-2), s. 31-68
- Kirsh, D.; Maglio, P. (1994): *On distinguishing epistemic from pragmatic action*. In *Cognitive Science*
- Kirsh, D. (1990): *When is Information Explicitly Represented?* In *Cognitive Science*, s. 340-365
- Klimeš, J. (2001): *Způsoby sledování pohybu zraku*. In *Strategie*, 9
- Komárek, S. (2000): *Příroda a kultura. Svět jevů a svět interpretací*. Vesmír, Praha
- Kosko, B. (1994): *Fuzzy Thinking. The New Science of Fuzzy Logic*. Flamingo, London
- Kosko, B. (1986): *Fuzzy cognitive maps*. In *International Journal of Man-Machine Studies*. 24, 1, s. 65-76.
- Kostroň, L. (1997): *Psychologie vytváření úsudků (teorie a metodologie Egona Brunswika, K. R. Hammonda a jejich následovníků)*. Masarykova univerzita, Brno
- Koukolík, F. (2006): *Sociální mozek*. Karolinum, Praha
- Koukolík, F. (2005): *Mozek a jeho duše*. Galén, Praha
- Koukolík, F. (2002): *Lidský mozek. Funkční systémy. Norma a poruchy*. Portál, Praha
- Kruskal, J.; Wish, M. (1978). *Multidimensional Scaling*. Sage, Beverly Hills, CA
- Kuhl, P. K.; Miller, J. D. (1975): *Speech perception by the chinchilla: Voiced–voiceless distinction in alveolar plosive consonants*. In *Science*, 190, s. 69–72
- Laird, J. E.; Newell, A.; Rosenbloom, P. S. (1987): *SOAR: An Architecture for General Intelligence*. In *Artificial Intelligence*. 33, s. 1-64
- Lakoff, G.; Núñez, R. F. (2000): *Where Mathematics Comes From. How the Embodied Mind Brings Mathematics Into Being*. Basic Books, New York
- Landau, B.; Jones, S.; Smith, L. (1992): *Perception, ontology, and naming in young children: Commentary on Soja, Carey, and Spelke*. In *Cognition*, 43, s. 85–91
- Langmeier, J.; Krejčířová, D. (1998): *Vývojová psychologie*. Grada Publishing, Praha
- Lave, J. (1988): *Cognition in Practice*. Cambridge University Press, Cambridge, MA
- Leibniz, G. W. (1932): *Nové úvahy o lidské soudnosti*. Bratislava

- Lewis, J. D.; Elman, J. L. (2002): *Learnability and the Statistical Structure of Language: Poverty of Stimulus Arguments Revisited*. In *Proceedings of the 26th Annual Boston University Conference on Language Development*. Cascadilla Press, Somerville, MA
- Liberman, A. M.; Harris, K. S.; Hoffman, H. S.; Griffith, B. C. (1957): *The discrimination of speech sounds within and across phoneme boundaries*. In *Journal of Experimental Psychology*, 54, s. 358–368
- Lloyd, D. (2004): *Radiant Cool. A Novel Theory of Consciousness*. MIT Press, Cambridge, MA
- Lloyd, D. (1989): *Simple Minds*. MIT Press, Cambridge, MA
- Locke, J. (1984): *Esej o lidském rozumu*. Svoboda, Praha
- Lukavský, J. (2005): *Sledování očních pohybů*. Bakalářská práce, nevydáno, Katedra softwarového inženýrství MFF UK, Praha
- Markman, E. M.; Wachtel, G. F. (1988): *Children's use of mutual exclusivity to constrain the meaning of words*. In *Cognitive Psychology*, 20, s. 121–157
- Markman, E. M.; Hutchinson, J. E. (1984): *Children's sensitivity to constraints in word meaning: Taxonomic versus thematic relations*. In *Cognitive Psychology*, 16, s. 1–27
- Marr, D. (1982): *Vision*. Freeman, San Francisco
- Mařík, V.; Štěpánková, O.; Lažanský, J. a kol. (2001): *Umělá inteligence (3)*. Academia, Praha
- Mattys, S. L.; Jusczyk, P. W.; Luce, P. A.; Morgan, J. L. (1999): *Phonotactic and prosodic effects on word segmentation in infants*. In *Cognitive Psychology*, 38, s. 465–494
- McClelland, J. L. (1989): *Parallel distributed processing: Implications for cognition and development*. In Morris, R. G. M. (Ed.): *Parallel Distributed Processing: Implications for Psychology and Neurobiology*. Clarendon Press, Oxford, s. 9–45
- McClelland, J. L.; Rumelhart, D. E.; Hinton, G. E. (1986): *The Appeal of Parallel Distributed Processing*. In Rumelhart, D. a kol. (Eds.): *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume I*. MIT Press, Cambridge, MA
- McLeod, P.; Plunkett, K.; Rolls, E. T. (1998): *Introduction to Connectionist Modelling of Cognitive Processes*. Oxford University Press, New York
- McLeod, P.; Dienes, Z. (1996): *Do Fielders Know Where to Go to Catch the Ball or Only How to Get There?* In *Journal of Experimental Psychology: Human Perception and Performance*, 22, s. 531–543
- Mehler, J.; Jusczyk, P.; Lambertz, G.; Halsted, N.; Bertoncini, J.; Amiel-Tison, C. (1988): *A precursor of language acquisition in young infants*. In *Cognition*, 29, s. 143–178

- Meier, R. P. (1991): *Language acquisition by deaf children*. In *American Scientist*, 79, 60-76
- Meltzoff, A. N.; Moore, M. K. (1983): *Newborn infants imitate adult facial gestures*. In *Child Development*, 54, s. 702-709
- Miall, C.; Tchalenko, J. (1998): *The Painter's Eye Movements*. [online] [cit. 12-3-2007] Dostupné z WWW <<http://www.physiol.ox.ac.uk/~rcm/pem/pem4.html>>
- Mikšík, O. (2003): *Psychologické teorie osobnosti*. Karolinum, Praha
- Mildeová, S; Vojtko, V. (2003): *Systémová dynamika*. Oeconomica, Praha
- Minsky, M. (1985): *The Society of Mind*. Simon & Schuster, New York
- Mitchell, M. (2006): *Complex Systems: Network Thinking*. In *Artificial Intelligence*, 170(18), s. 1194-1212.
- Mitchell, M. (1998): *Computation in cellular automata: A selected review*. In Gramss, T.; Bornholdt, S.; Gross, M.; Mitchell, M.; Pellizzari, T. (Eds.): *Nonstandard Computation*. VCH Verlagsgesellschaft, Weinheim, s. 95-140
- Mitchell, M.; Forrest, S. (1994): *Genetic Algorithms and Artificial Life*. In *Artificial Life*, 1 (3), s. 267-289
- Mitchell, M.; Hofstadter, D. R. (1991): *The Emergence of Understanding in a Computer Model of Concepts and Analogy-making*. In Forrest, S. (Ed.): *Emergent Computation*. MIT Press, Cambridge, MA
- Mls, K. (2005): *Fuzzy kognitivní mapy*. [online] [cit. 20-3-2007] Dostupné z WWW: <<http://lide.uhk.cz/home/fim/ucitel/mlska1/www/SOCO/prezentace%20Mls2005.ppt>>
- Nakisa, R. C.; Plunkett, K. (1998): *Evolution of a rapidly learned representation for speech*. In *Language and Cognitive Processes*, 13(2/3), s. 105–127
- Nakonečný, M. (1999): *Sociální psychologie*. Academia, Praha
- Newell, A.; Rosenbloom, P. S.; Laird, J. A. (1989): *Symbolic architecture for cognition*. In Posner, M. (Ed.): *Foundations of Cognitive Science*. MIT Press, Cambridge, MA
- Newell, A.; Simon, H. (1976): *Computer Science as Empirical Enquiring*. In Haugeland, J. (Ed.): *Mind Design*. MIT Press, Cambridge, MA
- Newsome, M. R.; Juszyk, P. W. (1995): *Do infants use stress as a cue in segmenting fluent speech?* In *Proceedings of the 19th Boston University Conference on Language Development*. Volume 2. Cascadilla Press, Somerville, MA, s. 415-426
- Norman, D. A. (1993): *Things that makes us smart: defending human attributes in the age of the machine*. Basic Books, New York
- Novák, V. (2000): *Základy fuzzy modelování*. BEN – technická literatura, Praha
- Novák, M. a kol. (1998): *Umělé neuronové sítě. Teorie a aplikace*. C. H. BECK, Praha

- O'Regan, J. K. (1992): *Solving the 'Real' Mysteries of Visual Perception: The World as an Outside memory*. In *Canadian Journal of Psychology*, 46, s. 461-488
- O'Reilly, R.C., Noelle, D., Braver, T.S.; Cohen, J.D. (2002): *Prefrontal cortex and dynamic categorization tasks: Representational organization and neuromodulatory control*. In *Cerebral Cortex*, 12, s. 246-257
- O'Reilly, R. C.; Munakata, Y. (2000): *Computational Explorations in Cognitive Neuroscience. Understanding the Mind by Simulating the Brain*. MIT Press, Cambridge, MA
- Papoušek, H.; Papoušek, M. (1987): *Intuitive parenting: a didactic counterpart to the infant's precocity in integrative capacities*. In Osofsky, J. D. (Ed.): *Handbook of infant development*. John Wiley, New York
- Pauková, E. (2005): *Aplikácia modelovania dynamických systémov pri riešení problematiky šírenia strachu z teroristickej hrozby*. Práce SVOČ, nevydáno, Katedra psychológie, FF UK, Bratislava
- Peregrin, J. (2005): *Kapitoly z analytické filosofie*. Filosofia, Praha
- Peregrin, J. (2003): *Člověk, pro kterého zítra již znamenalo včera*. In *Computerworld*, 14, s. 26
- Peregrin, J. (1999): *Význam a struktura*. Oikoymenh, Praha
- Peregrin, J. (1998): *Obrat k jazyku: Druhé kolo*. Filosofia, Praha
- Piaget, J. (1999): *Psychologie intelligence*. Portál, Praha
- Piaget, J.; Inhelderová, B. (1970): *Psychologie dítěte*. SPN, Praha
- Pinker, S. (2003): *Slová a pravidlá. Složky jazyka*. Kalligram, Bratislava
- Pinker, S. (1999): *How The Mind Works*. Penguin Group, New York
- Pinker, S. (1995): *Language acquisition*. In Gleitman, L. R.; Liberman, M. (Eds.): *Language: An invitation to cognitive science*. MIT Press & Bradford Books, Cambridge, MA
- Pinker, S. (1994): *The Language Instinct. The New Science of Language and Mind*. Penguin Group, New York
- Pinker, S.; Prince, A. (1988): *On language and connectionism: Analysis of a parallel distributed processing model of language acquisition*. In *Cognition*, 28, s. 73-193
- Pinker, S. (1987): *The bootstrapping problem in language acquisition*. In MacWhinney, B. (Ed.): *Mechanisms of language acquisition*. Erlbaum, Hillsdale, NJ, s. 399-441
- Pinker, S. (1984): *Language learnability and language development*. Harvard University Press, Cambridge, MA
- Plunkett, K.; Elman, J. L. (1997): *Exercises in Rethinking Innateness. A Handbook for Connectionist Simulations*. MIT Press, Cambridge, MA

- Plunkett, K.; Marchman, V. (1996): *Learning from a connectionist model of the acquisition of the English past tense*. In *Cognition*, 61, s. 299-308
- Plunkett, K.; Marchman, V. (1993): *From rote learning to system building: Acquiring verb morphology in children and connectionist nets*. In *Cognition*, 48, s. 21-69
- Plunkett, K.; Marchman, V. (1991): *U-shaped learning and frequency effects in a multi-layered perceptron: Implications for child language acquisition*. In *Cognition*, 38, s. 43-102
- Pylyshyn, Z. (1989): *Computing in Cognitive Science*. In Posner, M. (Ed.): *Foundations of Cognitive Science*. MIT Press & Bradford Book, Cambridge, MA
- Pylyshyn, Z. (1986): *Computation and Cognition*. MIT Press, Cambridge, MA
- Rádl, E. (1999): *Dějiny filosofie II*. Votobia, Praha
- Rains, G. D. (2001): *Principles of Human Neuropsychology*. McGraw-Hill, New York.
- Rorty, R. (2000): *Filozofia a zrkadlo prírody*. Kalligram, Bratislava
- Rosenbloom, P. S.; Laird, J.; Newell, A.; McCarl, R. (1992): *A preliminary analysis of the SOAR architecture as a basis for general intelligence*. In Kirsh, D. (Ed.): *Foundations of Artificial Intelligence*. MIT Press, Cambridge, MA
- Rosina, J. a kol. (1963): *Obecná psychologie pro pedagogické instituty*. SPN, Praha
- Rumelhart, D.; Smolensky, P.; McClelland, J.; Hinton, G. (1986a): *Schemata and Sequential Thought Processes in PDP Models*. In Rumelhart, D.; McClelland, J. (Eds.): *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 2*. MIT Press, Cambridge, MA
- Rumelhart, D.; Hinton, G.; Williams, R. (1986b): *Learning Internal Representations by Error Propagation*. In Rumelhart, D.; McClelland, J. (Eds.): *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1*. MIT Press, Cambridge, MA
- Rumelhart, D.; McClelland, J. (1986c): *PDP Models and General Issues in Cognitive Science*. In Rumelhart, D.; McClelland, J. (Eds.): *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume I*. MIT Press, Cambridge, MA
- Rumelhart, D.; McClelland, J. (1986d): *On learning the past tenses of English verbs*. In Rumelhart, D.; McClelland, J. (Eds.): *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 2*. MIT Press, Cambridge, MA
- Rumelhart, D. (1979): *Some problems with the notion that words have literal meanings*. In Ortony, A. (Ed.): *Metaphor and Thought*. Cambridge University Press, Cambridge, MA, s. 71–82
- Saffran, J. R.; Aslin, R. N.; Newport, E. L. (1996): *Statistical learning by 8-month-old infants*. In *Science*, 274, s. 1926-1928

- Santelmann, L.; Juszyk, P. (1998): *Sensitivity to discontinuous dependencies in language learners: Evidence for limitations in processing space*. In *Cognition*, 69, s. 105-134
- Sedláková, M. (2004): *Vybrané kapitoly z kognitivní psychologie. Mentální reprezentace a mentální modely*. Grada, Praha
- Seife, Ch. (2005): *Nula. Životopis jedné nebezpečné myšlenky*. Argo & Dokořán, Praha
- Siegelmann, H. T.; Sontag, E. D. (1995): *On the computational power of neural nets*. In *Journal of Computer and System Sciences*, 50, s. 132-150
- Siegler, R. (1981): *Developmental sequences within and between concepts*. In *Monographs of the Society for Research in Child Development*, 46 (2)
- Simon, H. A. (1991): *Models of my life*. Basic Books, New York
- Simon, H. A.; Kaplan, C. A. (1989): *Foundations of cognitive science*. In Posner, M. (Ed.): *Foundations of Cognitive Science*. MIT Press, Cambridge, MA
- Simon, H. A. (1982): *Models of bounded rationality*. MIT Press, Cambridge, MA
- Simon, H. A. (1981): *The Sciences of the Artificial*. MIT Press, Cambridge, MA
- Simon, H. A. (1979): *Information processing models of cognition*. In: *Annual Review of Psychology*, 30, 363-396
- Simon, H.; Newell, A. (1972): *Human Problem Solving*. Prentice-Hall, Englewood Cliffs, N.J.
- Simons, D. J.; Levin, D. T. (1997): *Change Blindness*. In *Trends in Cognitive Science*, 1, s. 261-267
- Singh, S. (2003): *Kniha kódů a šifer. Tajná komunikace od starého Egypta po kvantovou kryptografii*. Dokořán & Argo, Praha
- Skinner, B. F. (1957): *Verbal behaviour*. Appleton Century Crofts, New York
- Smith, L. B.; Thelen, E. (2003): *Development as a dynamic system*. In *Trends in Cognitive Science*, 7, s. 343-348
- Smolensky, P.; Legendre, G. (2006): *The Harmonic Mind: From Neural Computation To Optimality-Theoretic Grammar*. MIT Press, Cambridge, MA
- Smolensky, P. (1988): *On the proper treatment of connectionism*. In *The Behavioral and Brain Sciences*, 11, s. 1-74.
- Smolensky, P. (1987): *Connectionist AI, and the brain*. In *Artificial Intelligence Review*, 1, s. 95-109.
- Smolík, F. (2006): *Časná znalost jazyka: Vývoj receptivní znalosti jazyka v prvních dvou letech života*. In *Československá psychologie*, 50 (3), s. 238-250

- Smolík, F. (1999): *Psychologické aspekty osvojování jazyka*. Diplomová práce, nevydáno, Katedra psychologie FF UK, Praha
- Soderstrom, M.; Wexler, K.; Jusczyk, P. (2002): *English-learning toddlers' sensitivity to agreement morphology in receptive grammar*. In *Proceedings of the 24th Boston University Conference on Language Development. Volume 2*. Cascadilla Press, Somerville, MA, s. 643-652
- Soderstrom, M.; Jusczyk, P.; Kemler Nelson, D. (2000): *Evidence for use of phrasal packaging by English-learning 9-month-olds*. In *Proceedings of the 24th Boston University Conference on Language Development. Volume 2*. Cascadilla Press, Somerville, Mass., s. 708-718
- Sousedík, P. (1999): *Logika pro studenty humanitních oborů*. Vyšehrad, Praha
- Spelke, E. S. (1988): *Where perceiving ends and thinking begins: The apprehension of objects in infancy*. In Jonas, A. (Ed.): *Perceptual Development in Infancy*. Erlbaum, Hillsdale, NJ
- Spelke, E. S. (1990): *Principles of object perception*. In *Cognitive Science*, 14, s. 24-56
- Stehlík, L. (2007): *Nástroje mapování a modelování poznatkových schémat*. In Krejčová, L.; Mertin, V. (Eds.): *Studentská vědecká odborná činnost. Sborník prací z let 2005 a 2006*. Katedra psychologie FF UK, Praha
- Stephenson, N. (2006): *Kryptonikon*. TALPRESS, Praha
- Sternberg, R. J. (2002): *Kognitivní psychologie*. Portál, Praha
- Störig, H. J. (1993): *Malé dějiny filozofie*. ZVON, Praha
- Šlechta, P. (1998): *Kognitivní fuzzy systémy a jejich modelování*. Diplomová práce, nevydáno, Katedra psychologie FF UK, Praha
- Šmilauer, V. (1966): *Novočeská skladba*. SPN, Praha
- Šusta, M.; Neumaierová, I. (2004): *Cvičení ze systémové dynamiky*. Oeconomica, Praha
- Thagard, P. (2001): *Úvod do kognitivní vědy. Mysl a myšlení*. Portál, Praha
- Thelen, E. a kol. (2001): *The dynamics of embodiment: a field theory of infant perseverative reaching*. In *Behavior Brain Sciences*. 24, s. 1–86
- Thompson, R. K. R.; Oden, D. L.; Boysen, S. T. (1997): *Language-naive Chimpanzees judge relations between relations in an abstract mapping task*. In *Journal of Experimental Psychology: Animal Behavior Processes*. 23, s. 31-43
- Tincoff, R.; Santelmann, L.; Jusczyk, P. (2000): *Auxiliary verb learning and 18-month-olds' acquisition of morphological relationships*. In *Proceedings of the 24th Boston University Conference on Language Development. Volume 2*. Cascadilla Press, Somerville, MA, s. 726-737

- Výrost, J.; Slaměník, I. (Eds.) (1997): *Sociální psychologie – Sociálna psychológia*. ISV, Praha
- Watson, J. B. (1930): *Behaviorism*. Norton, New York
- Werker, J. F.; Tess, R. C. (1984): *Cross-language speech perception: Evidence for perceptual reorganization during the first year of life*. In *Infant Behavior and Development*, 7, s. 49-63
- White, R. (1994): *Jak pracují počítače*. UNIS publishing, Brno
- Wittgenstein, L. (1993): *Filosofická zkoumání*. Filosofický ústav AV ČR, Praha
- Wundt, W. (1900): *Die Sprache*. Englemann, Leipzig
- Wynn, K. (1995): *Evidence against empiricist accounts of the origins numerical knowledge*. In Goldman, A. (Ed.): *Readings in Philosophy and Cognitive Science*. MIT Press, Cambridge, MA
- Yarbus, A. (1967): *Eye Movements and Vision*. Plenum Press, New York
- Zadeh, L. A. (1973): *Outline of a new approach to the analysis of complex systems and decision processes*. In *IEEE Transactions on Systems, Man, and Cybernetics*. 1, s. 28-44.
- Zhang, J.; Norman, D. A. (1994): *Representations in distributed cognitive tasks*. In *Cognitive Science*, 18, s. 87-122
- Zrzavý, J.; Storch, D.; Mihulka, S. (2004): *Jak se dělá evoluce. Od sobeckého genu k rozmanitosti života*. Paseka, Praha

Internetové zdroje

- <<http://web.cecs.pdx.edu/~mm/>>
- <<http://www.cs.pomona.edu/~marshall/metacat/>>
- <<http://education.mit.edu/starlogo/>>
- <<http://www.bcp.psych.ualberta.ca/~mike/>>
- <<http://www.proverbs.cz/>>
- <<http://www.powersim.com>>
- <<http://www.ocnikamera.cz/>>
- <<http://viscog.beckman.uiuc.edu/grafs/demos/12.html/>>
- <<http://www.ochoadeaspuru.com/fuzcogmap/middleeast.php/>>
- <<http://hci.ucsd.edu/102a/05-lectures/W3a.pdf/>>
- <<http://crl.ucsd.edu/innate/tlearn.html/>>
- <<http://www.cnbc.cmu.edu/Resources/PDP++//PDP++.html/>>

Seznam doprovodného materiálu na přiloženém CD

Adresář *2D_CA*

Adresář *Metacat*

Adresář *PDP*

Adresář *StarLogo*

Adresář *Tlearn*

Adresář *Slepota_ke_změnám*

Adresář *Záznamy_z_oční_kamery*

Adresář *Hlásky_Exp*

Soubor *Turingův_stroj*

Soubor *Fuzzy_Expertní_System_Motivace*

Soubor *UNS_GA_XOR*

Soubor *FCM_Doprava*

Soubor *FCM_Doprava_GA*

Soubor *OptWorks_PC_Demo*

Soubor *Abakus*

Soubor *Log_pr*

Soubor *Lightavoid*

Soubor *Lightfollow*

Příloha

Rovnice modelu dynamiky strachu vytvořeného v simulačním prostředí Powersim Studio (převzato z www.proverbs.cz)

```
mainmodel dav {
const Frekvence diskrétních událostí {
autotype Real
init 0
permanent
}
aux Míra zvyšování úzkosti {
autotype Real
def GRAPH(Strach;0;10;{0,12;0,12;0,13;0,15;0,17;0,25;0,27;0,35;0,47;0,84;1//Min:0;Max:1;Zoom//})
}
aux Míra poklesu {
autotype Real
def GRAPH(Strach;0;10;{0,32;0,25;0,22;0,18;0,15;0,14;0,11;0,09;0,03;0,03;0,02//Min:0;Max:1;Zoom//})
}
aux Míra zapomínání {
autotype Real
def
GRAPH(Paměť_znepokojivých_událostí;0;40;{1,82;1,74;1,68;1,6;1,45;1,29;1,09;0,92;0,72;0,5;0,25//Min:0;Max:2;Zoom//})
}
aux Nárůst díky událostem {
autotype Real
def Příliv_znepokojivých_událostí*'Strach za událost'
}
aux Nárůst díky úzkosti {
autotype Real
def Paměť_znepokojivých_událostí*'Míra zvyšování úzkosti'
}
level Paměť_znepokojivých_událostí {
autotype Real
init 0
outflow { autodef Zapomínání }
inflow { autodef Příliv_znepokojivých_událostí }
}
aux Pokles_strachu {
autotype Real
def Strach*Míra_poklesu
}
aux Příliv_znepokojivých_událostí {
autotype Real
def IF (Spojitý_přiliv=1;Závažnost_spojitych;PULSE('Závažnost diskrétních událostí'; 3; 'Frekvence diskrétních událostí'))
}
const Spojitý_přiliv {
autotype Real
init 0
permanent
}
level Strach {
autotype Real
init 0
outflow { autodef Pokles_strachu }
inflow { autodef 'Nárůst díky úzkosti' }
inflow { autodef 'Nárůst díky událostem' }
}
Page 1
}
aux Strach za událost {
autotype Real
def GRAPH(Strach;0;10;{0,37;0,38;0,39;0,41;0,44;0,46;0,49;0,62;0,76;1,06;1,97//Min:0;Max:2;Zoom//})
}
aux Zapomínání {
autotype Real
def Paměť_znepokojivých_událostí*Míra_zapomínání
```



```
}  
const Závažnost diskrétních události {  
  autotype Real  
  init 0  
  permanent  
}  
const Závažnost_spojitéch {  
  autotype Real  
  init 0  
  permanent  
}  
}
```